PB168791

ENVIRONMENTAL HEALTH SERIES
Air Pollution
and
Water Supply
and Pollution Control

SYMPOSIUM

ENVIRONMENTAL MEASUREMENTS

Valid Data and Logical Interpretation

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND
TECHNICAL INFORMATION
Hardcopy | Microfiche | 309 pp | ao
$7.00 | $1.50
ARCHIVE COPY
Code 41

U. S. DEPARTMENT OF HEALTH
EDUCATION, AND WELFARE
Public Health Service

ENVIRONMENTAL MEASUREMENTS 999-AP-15 (or 999-WP-15)

SYMPOSIUM

ENVIRONMENTAL MEASUREMENTS

Valid Data and
Logical Interpretation

Sponsored by
Division of Air Pollution
and
Division of Water Supply and Pollution Control
September 4-6, 1963

Co-chairmen

JOHN S. NADER
Laboratory of Engineering and Physical Sciences, DAP
and
E. C. TSIVOGLOU
Technical Services Branch, DWS&PC

ROBERT A. TAFT SANITARY ENGINEERING CENTER
Cincinnati, Ohio

U. S. DEPARTMENT OF HEALTH
EDUCATION, AND WELFARE
Public Health Service

July 1964

The ENVIRONMENTAL HEALTH SERIES of reports was established to report the results of scientific and engineering studies of man's environment: The community, whether urban, suburban, or rural, where he lives, works, and plays; the air, water, and earth he uses and re-uses; and the wastes he produces and must dispose of in a way that preserves these natural resources. This SERIES of reports provides for professional users a central source of information on the intramural research activities of Divisions and Centers within the Public Health Service, and on their cooperative activities with state and local agencies, research institutions and industrial organizations. The general subject area of each report is indicated by the two letters that appear in the publication number; the indicators are

AP — Air Pollution
WP — Water Supply and Pollution Control
AH — Arctic Health
EE — Environmental Engineering
FP — Food Protection
OH — Occupational Health
RH — Radiological Health

Triplicate tear-out abstract cards are provided with reports in the SERIES to facilitate information retrieval. Space is provided on the cards for the user's accession number and key words.

Public Health Service Publication No. 999-AP-15
(or No. 999-WP-15)

PREFACE

The rapid development of air and water quality management programs within the Public Health Service and elsewhere has brought into sharper focus the many complex problems involved in obtaining valid environmental data from which to draw the most useful and valid conclusions. The availability of continuous measurement and recording devices, as well as the electronic computer, has made it possible to attempt the solution of increasingly complex environmental health problems that are associated with our expanding modern society. Complexity for its own sake is not a useful goal, however, and before we embrace the newer complex measurement and computational schemes we should take stock by deciding what it is we really wish to accomplish. Only thus can we rationally select the most suitable measurement system for a specific problem.

Although the problems associated with measurement systems are not unique to the environmental health field, some of the current needs of the Division of Air Pollution and the Division of Water Supply and Pollution Control of the Public Health Service led to this Symposium on Environmental Measurements. The Symposium Committee, in considering how best to approach the total problem, found it most susceptible to analysis by isolating each major operational step in the measurement system: sampling, detecting, recording, validating, interpreting, and drawing conclusions. This classification of operational steps provided the basic topics for General Sessions that would lead to better understanding of the operations common to diverse applications in environmental fields. We hoped in particular, by the very arrangement of the Symposium program, to emphasize that no measurement system can be any better than the weakest of the operational steps. Separate afternoon sessions were designed to explore the specific application of the operational steps to investigations of air and water environments.

It was hoped that this program orientation would enhance our understanding of the whole task of conducting a measurements program, and that it would thereby benefit pollution control and technical administrators, as well as researchers and scientists in the environmental field. It is anticipated that, following this orientation, there may well result a series of symposia, held from time to time, to discuss in more depth specific operational aspects of measurement programs. The purpose of the Symposium, therefore, was to provide comprehensive orientation; it was intended more to raise questions than to provide solutions.

THE CHAIRMEN

SYMPOSIUM COMMITTEE

Division of Air Pollution	**Division of Water Supply and Pollution Control**
CO-CHAIRMEN	
John S. Nader Laboratory of Engineering and Physical Sciences	Dr. E. C. Tsivoglou Technical Services Branch
COMMITTEE MEMBERS	
Robert A. McCormick Laboratory of Engineering and Physical Sciences	Richard O'Connell Technical Services Branch
Charles E. Zimmer Laboratory of Engineering and Physical Sciences	Alfred W. Hoadley Basic Data Branch

CONTENTS

Page

Welcome: J. E. Flanagan 1

Session 1 — General

C. S. Draper: Information Engineering — New Frontier of Technology 5
G. W. Anderson: Objectives of Measurement Systems 11

Session 2 — General

P. W. MacCready: The Design of Measurement Systems 21
A. Goetz: Parameters 29
W. J. Youden: Sampling and Statistical Design 35
L. Bollinger: Transducers 41

Session 3 — General

P. K. Stein: Classification Systems for Transducers and Measuring Systems 65
G. C. Gill: Data Validation 85
R. S. Green: The Storage and Retrieval of Data for Water Quality Control — a Summary 101

Session 4 — Measurements of Air Environment

J. S. Nader: Data Acquisition Systems in Air Quality 107
H. E. Cramer: Data Acquisition Systems in Meteorology 125
O. Balchum: Data Acquisition Systems in Physiology 140

Discussion: Data Acquisition Systems 159

Session 5 — Measurements of Water Environment

S. S. Baxter: Data Acquisition Systems in Water Supply 163
P. DeFalco, Jr.: Data Acquisition Systems in Water Quality Control 173
W. Isherwood: Data Acquisition Systems in Hydrology 179
J. J. Gannon: The Interpretation and Analysis of Hydrological Data 187

Session 6 — General

J. C. Bellamy: Data Display for Analysis 213
G. W. Brier: Techniques for Data Analysis 227
D. W. Pritchard: Interpretations and Conclusions 235
L. A. Chambers: Summation 245

Session 7 — Measurements of Air Environment

R. I. Larsen: Determining Basic Relationships between Variables 251
G. W. Brier: Interpretation of Trends and Cycles 265
L. D. Zeidberg: Data Interpretation (Air) — Drawing Conclusions 273

Discussion: Interpretations and Conclusions 285

Session 8 — Measurements of Water Environment

H. B. N. Hynes: The Interpretation of Biological Data with Reference to Water Quality 289
W. Stumm: Chemistry of Natural Waters in Relation to Water Quality 299
G. A. Rohlich: Data Interpretation (Water) — Drawing Conclusions 325

Joseph E. Flanagan, Jr.[*]
Acting Director
Robert A. Taft Sanitary Engineering Center

WELCOME

I think it is fair to say that the opportunity to open a session like this ranks very high among the very numerous and pleasant duties that fall to the Director of the Taft Sanitary Engineering Center. I am not going to give you a discourse on the Center. You can draw many conclusions about us just from this particular program. Since this is a center for multiple environmental health programs, it is particularly appropriate that we are able to sponsor a meeting of this type. I think that the most significant point about this session is that it is sponsored by two of our operating divisions, probably the first time that this has happened. This particular symposium is a direct outgrowth of communications and rather constant associations between scientists of the Air Division and of the Water Division. Representatives of our third, fourth, and fifth divisions are also present at the meeting today, and I trust that everyone will get a good bit from it.

One of the peculiar situations about opening meetings is that you are supposed to say a word of "welcome." I will never forget attending a session similar to this where one of the rather high-level administrators at a school stood on a platform like this and said, "I've been told that I'm supposed to welcome you. Well of course, you're welcome," he said, "but it just seems like an undue interference with the program for me to stand here and continue along this line."

I had the privilege this morning of having breakfast with the gentleman who is going to give the keynote address. About the only thing that I am going to say is that if the rest of this program stands up to what I think this keynote address is going to be, after meeting this chap and chatting with him informally, I think we are all due for a worthwhile experience. So, welcome to Cincinnati — nice to have you with us.

---

[*] Now Associate Director, Department of Environmental Health, American Medical Association, Chicago.

SESSION 1: General

Chairman: **Keith S. Krause**
Chief, Technical Services Branch
Division of Water Supply and Pollution Control
U. S. Public Health Service

Dr. Charles S. Draper
Head, Department of
Aeronautics and Astronautics
Director, Instrumentation Laboratory
Massachusetts Institute of Technology, Cambridge

SUMMARY

Information engineering is described as the region of human activity that deals professionally with the conception, design, building, testing, manufacture, and operation of components and systems to sense physical quantities and from these as inputs, to generate operating commands for the machines and the organizations that serve the needs and desires of mankind. Automatic information systems are now recognized as necessities throughout the realms of science, business, industry, and transportation. The evolution of this important frontier of today's technology is described here along with the various elements that contribute to such a system. A broad and accurate knowledge of the environment has become essential for the health, economic welfare, and general progress of the human race. From an engineering standpoint the transmission, processing, indicating, and recording signals that represent the environmental information have been established; the difficult problems remaining involve the balancing of benefits from results to be expected against funds and other resources that must be made available.

INFORMATION ENGINEERING — THE NEW FRONTIER OF TECHNOLOGY

Information engineering is described as the region of human activity that deals professionally with the conception, design, building, testing, manufacture, and operation of components and systems to sense physical quantities and from these as inputs, to generate operating commands for the machines and the organizations that serve the needs and desires of mankind. The devices that provide for these purposes by sensing, transmitting, processing, and applying information are called instruments. The complex of instruments used to meet the information-handling requirements of a particular set of circumstances is the instrumentation for the given situation. The over-all technology of instrumentation is the sum total of knowledge, engineering, devices, resources, facilities, manpower, and services that are directed toward the realization of means to fulfill the information requirements of civilization.

Instrumentation is based on components and subsystems designed for sensing, communicating, processing, and using information. Today, the complex of industry and business devoted to providing instrumentation for the United States involves operations at the billion-dollar-per-year level; a level that has grown by several orders of magnitude during the last 4 decades. This expansion has naturally been accompanied by the addition of many thousands of scientists, engineers, and technicians to the ranks of those who are concerned with instrumentation as a profession. The end is not yet — even if we recognize the revolutionary growth that has already occurred, instrumentation is still so far from exhausting requirements and possibilities that it must be classed among the pioneering areas of human activity. Because of its universal importance for other areas and its very great remaining growth potential, instrumentation may reasonably be considered as a most important frontier region of today's technology.

This special situation, which is associated with the means for handling information, does not come from any recent discoveries. Rather, the increasing importance of the field depends on an accelerating shift from human operators to inanimate equipment that provides revolutionary new features in capacity, speed, reliability, accuracy, size, weight, operating cost, and general utility. Recent improvements in these characteristics have surely expended the feasibility regions of instrumentation and brought into a clear focus some situations that were largely unrecognized until the middle years of this century. The situations in question generally follow the model of those occurring in a human body as it reacts to external stimuli. This response involves power level actions from bone and muscle structures as they follow the command signals generated by the brain on the basis of signals from the sense organs. These commands are sent over nerve paths to muscles that respond by actions suited to the situation represented by the sensor signals. The correctness of these responses is determined by the sensors and brain that work together in comparing actual positions and motions with intended positions and motions. Deviations of actual from desired conditions appear as feedback signals within the human information system. These error signals cause the command signals to change in ways that bring about corrective changes in the muscle actions. Behavior of this kind, which applies feedback information for control purposes, has been present in all high-level animals since the beginning of their existence. It is interesting to note that feedback control has only recently appeared in the governors, the controllers, and the servomechanisms that are now essential parts of substantially all operating systems.

Until a new era started some 200 years ago, the progress of mankind was primarily concerned with increasing the power of organizations and equipments. For example, in ancient times, ships became bigger and stronger with propulsion by larger sails and more oars. Such ships did not, however, become useful until their actions were controlled by orders from commanders and pilots. Regarded as parts of over-all systems, the human being in charge functioned by sensing information, combining it with essential facts and plans through reasoning processes and applying the resulting decisions to the power systems involved by means of command signals.

The powered effector subsystems for carrying out such commands have never existed in nature. It has always been necessary for men to conceive, design, and build effector systems. On the other hand, the *information subsystem*, always a necessary part of any *over-all system* and effectively the "mirror image" of the *power subsystem*, could depend on the senses, brain, and nerves of some human being to care for all its essential functions. For thousands of years, this availability of human information subsystems that could be easily matched to power subsystems much stronger than any individual prevented attention from being directed toward a clear realization of the essential role played by information in all the devices of technology. For example, from the beginning of navigation, ships were controlled by the judgment and skill of a single man. Armies of great size were commanded by general officers using disciplined men in a proper organization as the means of control. Perhaps communications were slow, but given enough time, military machines could usually be made to follow the plans of their accepted leaders.

It is true that ships were wrecked and armies were lost by failures of their information systems to cope with difficult situations, but on the whole human beings met the needs of such systems well enough until the age of mechanized power began in the eighteenth century. First, it was the substitution of automatic gear for the uninspiring job of manipulating steam valves on the basis of piston position in a pumping engine, then it was the use of a centrifugal governor to keep engine speed constant by changing steam flow, a chore that human operators could not have performed well in any case. Regulators for pressure, temperature, and voltage followed the automatic timing devices for engines, affording relief from simple repetitious tasks. Such devices became commonplace during the first half of the twentieth century, and are still being developed in the direction of more sophistication and higher performance.

This same time period saw, among other advances, the realization of complex electric power systems depending on accurate adjustment of voltage and frequency, the building of very large high-speed ships, the use of high-performance aircraft, the development of ballistic missiles, and the design of vehicles for travel through space. The requirements laid on information systems by these new devices forced the necessary performance well beyond the capabilities of human beings. The factors introducing difficulties include complexity, accuracy, speed of response, length of working periods, reliability, and environments too severe for comfort or in some cases even for survival of human beings. Modern supersonic aircraft still carry pilots but provide many radiation sensors, automatic adjustments, and booster devices to assist with information system functions. Ballistic missiles with their one-way missions and space ships with no men aboard actually force information system designers to use only inanimate elements. The firmly demonstrated fact that in-production self-contained guidance systems of this kind can receive, process, and apply information well enough to produce hits at great ranges is only one of many proofs that the era of the automatic information system is not only beginning but is already well on its way.

Evidence is all around us that information systems are now recognized as necessities throughout the realms of science, business, industry, and transportation. The worldwide credit card organizations and airline reservation services could not exist without very rapid, accurate, and reliable collection, transmission, and processing of information. At least one airline uses an information system in which an agent, in San Francisco for example, receiving a request for a reservation punches keys to send a signal across the country to a central computer in New York State. Results from this computer, in terms of signals sent over telephone lines, return information on seat availability within a few seconds. This example merely illustrates one of the ways that information systems are revolutionizing the operations of modern society. Only the surface has been scratched as yet; very wide regions remain to be explored and exploited by able individuals dedicated to the professional practice of information engineering. Today this field truly belongs to the frontiers of technology.

ELEMENTS OF INFORMATION ENGINEERING

Information engineering is concerned with applying scientific knowledge, professional education, experiences, judgment, initiative, and perseverance in the use of natural resources, facilities, available funds, and the capabilities of technology so that we can realize information systems and their components able to meet stated specifications within given limits. Briefly, the engineer undertakes to produce certain practical results under the restrictions of existing circumstances. His particular stock in trade is state-of-the-art technology, imagination concerning future developments, and the know-how to build components and techniques into satisfactorily working systems. In fact, he has many well-developed and extensive segments of technology at his disposol. These component technologies are concerned with devices and systems that fall into five principal classes:

1. *Sensors*, the devices that receive states of physical quantities as inputs and produce signals representing these states as outputs.

2. *Communication* systems that transmit signals among information subsystems.
3. *Coupling systems* that modify output signals from one subsystem so that these signals are suitable for inputs to other subsystems.
4. *Computing systems* that receive one or more independent signals as inputs, carry out logical operations, and produce outputs that represent information derived from the inputs.
5. *Display and recording* systems that provide direct visual indications and records of signals and the information with which these signals are associated.

Each of these five categories is now the basis for a more or less distinct area of over-all technology. Some of these component technologies are very large and broad within themselves, while others are smaller but still have magnitudes of some consequences in the world of industry.

For example, communication surely includes all the techniques of wired telegraphy, telephones, radio, and radar. Indeed, certain information systems may include transcontinental telephone lines, submarine cables, and million-mile radio links from earth to space vehicles. On the other hand, communication within a system may involve no more than a few inches of wire.

Couplers concern a relatively small number of specializing companies that deal with such components as amplifiers, transformers, digital-to-analogue converters, data storage systems, etc.

Display and recording devices have many forms, with analogue and digital presentations ranging from lines on high-speed cathode ray tubes and electroluminescent figures to typed numbers and inked points or curves on paper sheets. Equipment of this kind is available from a number of companies that are able to provide standard arrangements and to meet the needs of special situations.

Computing systems are the basis for a major industry supplying both analogue and digital computers that operate over a wide range of speeds, capacities, and complexities. It is certain that substantially any data-processing requirements of practical importance can be fulfilled by currently available techniques. The differences among various designs lie in capacity and speed for given weights and sizes, matters that have all received and are still receiving great attention. Very large computer installations exist today for solving complex problems, and very small units to deal with the complex but specialized situations associated with missles in flight are also in production. Performance is generally adequate, but the premiums for lighter, faster, and more reliable equipment are so great that developments can be expected to continue for a long time into the future.

Because electrical signals are especially suitable for representing any kind of data with low power levels, can be transmitted in many ways over short and long distances, and are easily adapted for rapid processing, electronic techniques are very widely applied in information system designs. Pneumatic, hydraulic, and mechanical principles may be and have been used in computers, but recent tremendous developments of electronic devices such as the widely used magnetic transistors and magnetic tape recorders will surely continue to force information systems toward electronics.

SENSORS

Sensors are instruments that respond to states of physical quantities as inputs and deliver representations of these states as their outputs. For example, a mercury-in-glass thermometer is a sensor for temperature as its essential input and produces the length of a mercury column index as its output. When *reference marks* are placed near the index so that "higher" and "lower" readings may be qualitatively distinguished, the sensor is an *indicator*. If a scale having figures related to a series of systematically placed reference marks is used to associate a number with each state of the input, the sensor becomes a *measuring instrument*. When the sensor output is a signal of a kind that has a series of states uniquely associated with corresponding states of the input, the sensor becomes a *signal generator*.

In practice, sensors may simultaneously serve each of the three output functions. A sensor may have an on and off light for the purposes of indication and a scale and pointer arrangement for measurement, and at the same time, may produce signals representing input states. Indications have various forms including index positions, "flag" exposures, lights, color gradations, etc. Measurements, by definition, always involve numbers, but output signals may have many forms. Continuously varying pressures, gas or liquid flow rates, current levels, voltage magnitudes, mechanical displacements, and other configurations are all used as sensor output signals. In recent years, discrete pulses of fluid or electricity have come to be widely used as signals. These digital signals have the great advantage that they can be used as direct inputs to modern computers and data-processing systems. It is to be expected that as time goes on all information systems will be based on such signals.

Sensors are often called *transducers* because their functions involve "transducing," that is, carrying power from one region to another region. The term "sensor" is preferred because it stresses the fact that information, not power, is the essential factor in the primary function of a sensor.

Of all the elements that make up information systems, sensors have the longest and most honorable background in history. Devices to measure physical quantities really helped very much to start modern science by providing sound experimental information to supplement, and often to replace, the pure philosophy that scholars had inherited from ancient Greece. Always, careful description of measuring equipment and proof of accuracy formed substantial parts of any scientific paper. For purposes of this kind each instrument was conceived and built to perform a specific task, usually one of laboratory measurement. There was no consistent pattern of concepts, terminology, or design. Outputs usually took the form of numbers read from a scale and index combination by observers who wrote down readings as they appeared. High accuracy for the slow changes and completely static condition was the performance objective, rather than ability to handle many inputs at high speed. Much attention was devoted to compensating out or correcting for errors caused by environmental effects. Very fine instruments were developed for research purposes. In general, these devices were not well suited for reliable operation under severe service conditions.

With the development of ships, stationary power plants, automotive vehicles, and aircraft, indicators and measuring instruments emerged from the laboratory and became integral working components of information systems. About 1930 the universal procedure of regarding theory and practice for each instrument type as a special isolated section of science began to be replaced by methods in which concepts and notation were made consistent and adapted to the purposes of information systems having arbitrary levels of complexity. This process has continued until the engineering of sensors is now a well-established part of system technology, with production components available that meet written specifications on the basis of routine inspection operations.

Certainly, it is now practical to engineer information systems in which the reception of data by sensors is a well-defined and reliable aspect of operation. A wide range of sensors is available from manufacturers who list them as catalogue items. When the requirements of a particular situation cannot be met by production instruments, the whole background of science and technology is available for use in designing special instruments.

To meet the needs of present-day information collection, research work, and control operations, multiple sensing units are formed into patterns that serve the purpose of making simultaneous observations over an extended field of physical quantities. The complex of worldwide meteorological stations is an example of a system that uses multiple sensors to cover pressure, temperature, humidity, wind velocity, etc., for wide geographical areas.

In general, indications, measurements, and signal generator outputs, all of low or high quality as required, are well provided for by existing sensors or by principles that can be embodied in practical instruments if this is necessary. The present frontier lies not in sensors but in systems for collecting, processing, and applying information from areas of significant size for understanding and control of essential conditions.

ENVIRONMENTAL INFORMATION SYSTEMS

Broad and accurate knowledge of the environment has become essential for the health, economic happiness, and general progress of the human race. Adequate information on the conditions in water supplies, lakes, rivers, and ocean shores is necessary for the prevention of illness. Data on watershed situations, rain and snow, are required for the control of floods by the operation of storage reservoir spillways. Atmospheric conditions and wind patterns over wide areas must be systematically known for weather prediction purposes. Smog from air pollution must be known if the public health is to be properly protected and undesirable conditions brought under control.

Problems of sensors to indicate, measure, and generate signals for collecting any amount of information on the environment are surely not difficult. From the standpoint of technology, the transmission, processing, indicating, and recording signals that represent the environmental information are all matters of established engineering. The interpretation of results in terms of safety and control measures may be subjects of some controversy, but should not hold up major decisions.

The really difficult problems involve the balancing of benefits from results to be expected against funds and other resources that must be made available. At the present time, information systems to sense and interpret environmental data in terms suitable for controlling reactions are just beginning to demonstrate their usefulness; however, it is only a question of time before complete coverage networks will send environmental information to central computers and display centers from which fast and effective decisions may be made on proper reactions to correct or control undesirable situations. The next few years will surely be exciting and interesting for both scientists and engineers who carry the responsibility of dealing with the human environment.

Dr. Gaylord W. Anderson
Director, School of Public Health
University of Minnesota, Minneapolis

SUMMARY

The basic reason for measuring the factors of man's environment is to determine the magnitude of the various external forces and, insofar as possible, the effect these forces have on man. Examples are cited that point out the difficulties encountered in establishing valid associations between environmental variables and human disease. Unfortunately, many of the measurements used to evaluate the magnitude of environmental hazards are no more reliable than are certain data on the occurrence of diseases that we attempt to attribute to these hazards. As we approach the problem of environmental measurements, our objectives must be twofold: We must seek ways of measuring the magnitude of a vast array of environmental variables that may conceivably have a bearing on human health, and we must attempt to measure the development of human disease so that we may correlate these findings with the results of environmental measurements.

OBJECTIVES OF MEASUREMENT SYSTEMS

The World Health Organization in its charter has defined public health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." If we accept this forward looking definition of public health, as have over 100 nations in joining the Organization, then is it obvious that in considering man's environment and attempting to establish valid bases for measurement we must concern ourselves with a vast array of factors that, through their effect on the environment, may adversely affect the physical, mental, and social well-being of mankind. The purpose of our conference is to attempt to determine to what extent and in what manner we may measure some of these factors and determine their significance for man.

Undoubtedly, we could find as many different definitions of environment as there are registrants at this conference, and no one of us could rightfully claim for his definition superiority over that of his colleagues. For my purposes and for the purposes of this discussion, I like to think of environment as the sum of all the external biological, chemical, and physical forces that surround man and therefore may influence his body processes or his behavior. Under this broad concept, you and I can be considered as part of each other's environment for we can spread infection to each other, or through our personalities or behavior can not only irritate one another but can at times actually jeopardize human life.

The usual connotation of the term sets aside the human factor, however, and concerns itself with the physical, chemical, and non-human biologic forces that surround man. Included in the latter are not only the pathogenic forms of animal and plant life but also those animals that serve as reservoirs of infection and those insects that serve as vectors. For the purposes of this conference, however, we may focus our attention on the physical and chemical factors of the environment and set aside the biologic and human elements. Yet, we cannot in truth set these aside for, even though we may limit our discussion to a consideration of measuring air, water, or ionizing radiation we must remember that to a very high degree it is man and his behavior that have introduced hazards into these media. Industrialization resulting from human discoveries has resulted in pollution of our environment. Economic forces have resulted in demands for products, the production of which results in environmental hazards not alone to the worker or the consumer, but also to the general public. Sociological and political forces emanating in part from technological developments have created suburbanization with the resultant creation of an entirely new set of environmental hazards. Cultural forces, expressed in human behavior, are in many parts of the world of fundamental importance in pollution of the environment, for example, the disposal of human excreta in such a manner as will endanger the water that man must drink or in which he must work. Thus, while we may, in a conference such as this, limit our discussions to the fairly specific problems of measuring some of the physical, chemical, or microbial hazards of our environment, we must never forget that behind these factors are vast economic, sociologic, cultural, and political forces that, in the last analysis, are responsible for endangering the environment and, in the years ahead, will in all probability add to this danger. To a certain degree, we who are concerned with the control of the environment must resemble the fabled King Canute, who is reputed to have attempted to sweep back the tide from the beaches with his broom. Unquestionably, our brooms of control, while less regal, will have more effect, yet we must never forget that these other forces are more fundamental, that they are part of an inevitable social evolution that is far less easily controlled or regulated than are the tangible components of our environment.

Keeping in mind this concept of environment, we may now ask ourselves why we wish to measure various factors of our environment, what factors we wish to measure, what conclusions we may wish to derive from these measurements, and finally, how we shall proceed with such measurements. This latter I shall leave to others far more competent mathematically than I can ever hope to be; yet from the point of view of an epidemiologist, I may have the temerity to make a few suggestions.

Our basic reason for measurement is to determine the magnitude of the various external forces in man's environment and, so far as possible, the effect that these forces have on man. Ideally, therefore, we are seeking to measure both cause and effect, variables that are interrelated, in that the causal force, if sufficiently great, may be harmful to man, who is exposed to the force. If this cause-and-effect relationship is known or can be established, we can then presume that the effect will increase or decrease to the extent that we are able to alter the magnitude of the cause, other secondary or related factors being equal and constant. Unfortunately, however, we can rarely assume that these other factors are static for the social and biologic relationships between man and his environment are conditioned by a vast array of changing variables. In the laboratories of physical science, it is often possible to study the effect of a single variable by stabilizing all other components of a reaction system; yet in the biologic world, and especially in the study of man in his normal community environment, we can rarely, if ever, keep all but one variable constant.

May I illustrate this point with a simple example from the realm of the infectious diseases, and notably typhoid fever, the control of which stands as a magnificent monument to environmental sanitation? The spread and development of any infectious disease may be considered as consisting of six components, the etiologic or causative agent, the reservoir of infection, the escape from the reservoir, the transmission to a new host, the entry into this host, and the susceptibility of the host. We may easily measure the extent of contamination of water, a well-known and easily performed measurement, and may also count the number of typhoid cases and deaths. It would be simple to correlate these two measurements and arrive at certain very satisfying conclusions, which might, however, be erroneous because such a correlation would overlook such variables as the size of the reservoir and the susceptibility of the host.

The accepted measurement of safety of water for human consumption has been the so-called Treasury Standards, going back to the era when the Public Health Service was a part of the Treasury Department. These standards are based on the presence or absence of gas-forming organisms in detectable quantity in various quantities of water, it being correctly assumed that these gas-formers are of intestinal origin. The standards are thus a measure of sewage pollution, not of contamination with typhoid organisms, the detection of which has been technically impossible with a satisfactory degree of accuracy. In the establishment of these standards and the acceptance of water that did not have more than a certain amount of demonstrable sewage pollution, there was an assumption that this same water did not have enough typhoid organisms to be dangerous to the consumer in the quantities he might reasonably be expected to drink. In other words, these standards were based on an assumed ratio between the numbers of typhoid and of colon bacilli in the sewage of a given community. If the sewage pollution did not exceed a certain level, we could assume that the number of typhoid bacilli was below the danger level. Experience showed this assumption to be correct.

Yet, as time has passed and as we have observed the trends in water pollution and typhoid incidence, we have been forced to recognize that these relationships no longer hold. The number of colon bacilli that an individual or a unit population contributes to sewage has remained essentially stationary, since we are dealing with organisms that are normal and invariable inhabitants of the human intestine. On the other hand, the reservoir of typhoid, chiefly in the form of carriers, has declined strikingly as the carriers have died off faster than they have been replaced. Thus, while the number of gas-formers per unit quantity of sewage may have remained unchanged, the number of typhoid organisms has declined and the older ratio between gas-formers and typhoid bacilli is no longer valid. In other words, so far as the risk of typhoid is concerned, we may safely drink water that contains far more colon organisms than could have been safely consumed 20 to 50 years ago. Our epidemiologic experience confirms this. In former days, sewage pollution of water resulting in a short-lived community-wide outbreak of diarrhea was invariably followed by typhoid. In recent years, we have had innumerable instances of community-wide diarrhea due to accidental sewage contamination of water; yet typhoid has not ensued, simply because the reservoir of typhoid carriers has dwindled to a point at which the number of typhoid bacilli in the sewage is too small to cause disease. Thus the old standard, while still defensible on aesthetic grounds as a matter of common decency and possibly still as a measure of other pathogens that may contaminate the water, is no longer a valid measurement of the safety of the water as a vehicle for the spread of typhoid, even though the public, through its continuing lack of exposure to typhoid, is more susceptible today than in former years.

This very change in susceptibility may, however, have an opposite effect in a situation in which the number of organisms available for ingestion may be no greater than in former years. As one examines the records of food-borne outbreaks of typhoid during the past half century, one notes that although the number of such outbreaks has declined the attack rate among those assembled for a meal prepared by a carrier has become progressively greater. The decline in number of outbreaks can be attributed to the reduction in the number of carriers while the increased attack rate is due to the fact that with the declining incidence of cases and prevalence of carriers fewer persons have been latently immunized through repeated small doses of organisms. Any standard that might have been developed to determine the number of typhoid organisms in food, even if accurate, would have afforded little clue to the human effect, unless at the same time we had taken cognizance of the susceptibility of those who were to eat the food. Thus, mere measurement of the number of typhoid bacilli in food would not have given a true picture of the hazard of consumption of the food.

Two more examples will suffice to point out some of the problems of environmental measurements as indicators of hazards to human health. It is a well substantiated fact that milk from cows infected with Q-fever contains large numbers of the causative rickettsiae. Equally well established is the failure of pasteurization carried out in accord with present standards to kill certain strains of the Q-fever organisms, with the result that persons drinking such milk not only can but actually do develop a rickettsial infection as manifested by the development of specific antibodies. On first thought, therefore, one might logically conclude that our criteria for satisfactory pasteurization should be changed to require a higher temperature so as to kill these organisms. Yet, in spite of incontrovertible evidence of infection as a result of drinking milk containing viable rickettsiae, there is no evidence that clinical illness has resulted. There are reasons for believing that although serious and even fatal illness may result from inhalation of very small numbers of the Q-fever organisms ingestion in even large amounts does not produce disease but only latent unimportant infection, which may actually be beneficial in that it may possibly immunize the individual against illness if at a later date organisms are inhaled. Measurement of the number of organisms in milk may therefore have little meaning or significance; yet great importance can be attached to any measurements of the number of organisms suspended in the air, their survival in the air, and the physical forces that govern their dispersion into the air. Quantitative studies of the production of aerosols may have tremendous significance in the understanding of this as well as of several other infectious diseases and of various conditions attributed to inhalation of chemical agents.

For a third example, I should like to turn to poliomyelitis, a situation in which we unfortunately find an amazing amount of unreliable data, both as to occurrence of the disease and causative environmental factors. Much has been published regarding trends in the incidence of the disease and innumerable attempts made to correlate these apparent trends with various forces that lend themselves to easy though not always highly accurate measurement. The sad fact is that we have few reliable statistics that can be used to determine trends of this infection. Morbidity or even mortality data are of little value, because diagnostic criteria have changed tremendously from the era when only the severely paralyzed were counted to a later period when the much more numerous non-paralytic infections were included in the report data but not clearly separated from the paralytic cases. The situation has become even more complex of recent years as we have come to recognize that a high proportion of the non-paralytic cases clinically diagnosed as poliomyelitis are in fact infections with other viruses, the true nature of the infection being determinable only by laboratory procedures that may be expensive of both time and personnel and hence are not routinely performed. Even more disturbing from the statistical standpoint is the fact that certain other viruses may produce paralytic conditions clinically indistinguishable from those of true poliomyelitis. Only in recent years have we had practical laboratory procedures that will identify the true etiology of such infections, and even today there is no universal use of such tests in diagnosis. Thus, even though we may make exact measurements of environmental factors such as improvement in various aspects of community sanitation, correlation of these findings with those of poliomyelitis incidence would be of little value because of the highly inaccurate nature of such incidence data.

This very problem may well be the basis for no small amount of controversy in the years immediately ahead of us. It is common knowledge that with the suburbanization of our large city population vast numbers of persons previously served by public water supplies and sewerage systems have moved to areas where reliance is placed on individual wells and septic tanks or cesspools. That the waste products may drain into the wells is amply shown by the amount of household detergent in the water from such wells.

Currently, we are seeing the widespread use of oral vaccines for immunization against poliomyelitis, vaccines containing living attenuated organisms that pass through the intestinal tract and are given off in the feces for variable periods of time. Although the survival of these in water and sewage has not been well studied, it is conceivable that they may appear in various water supplies. Could we infer therefore that by demonstration of such virus or by any measurement of its concentration in water we could establish a hazard that required the development of new control measures? I doubt it, since we would be dealing with a virus designed for human ingestion as an immunizing agent and therefore safe. In fact, were we to demonstrate that the excreted oral vaccine passed into the untreated water of such wells, we might conceivably look upon it as a valuable means of immunizing certain persons who had failed to take the vaccine for their own protection. All of this may not occur. I introduce the possibility merely to point up the need for proper interpretation of data, however exactly they may be determined.

Unfortunately, many of the measurements used to evaluate the magnitude of environmental hazards are no more reliable than are certain data on the occurrence of diseases that we attempt to attribute to these factors. I am reminded of a study, unfortunately published and made the basis of editorial comment, in a reputable medical journal that purported to show an inverse correlation between the hardness of water and the development of coronary disease. Since this correlation purportedly existed, the inevitable conclusion was advanced that softening of water was undesirable as it carried with it an adverse effect on the cardiovascular system of the user and might therefore be a factor in the high mortality rate from heart diseases. One is accustomed to accept cardiovascular morbidity and mortality with a certain degree of caution because of diagnostic difficulties; yet one can expect a reasonably exact measurement of the hardness of water. Yet, as one examined the data, one learned to his unbelieving astonishment that the author had taken for each state in the Union a single numerical value that was supposed to represent the hardness of water throughout the state. Not only had he ignored geological factors and the resultant great variations in the hardness of the public water supplies within a given state, but he had also ignored movements of population, assuming that the individual throughout his life had been under the influence of the hardness of the water of the state in which he drew his last breath. It is almost incredible to me that data so obviously unreliable should be published and made the basis of conclusions, both by a scientific journal and by such eminently reliable sources of medical news and opinion as our well-known popular news digests. Even more amazing and naive is the statement in the article that the data have superior reliability and significance because they have been processed by an electronic computer.

I realize full well that in my attempt to point out some of the difficulties in establishing valid associations between environmental variables and human disease I may have appeared to be purely destructive. If such be the illusion that I have created, I must beg your forgiveness; although I recognize the unavoidable difficulties, I equally recognize the importance of such correlations whenever they are valid. In all too many instances, we are presented with environmental variables, the significance of which is still problematical. We can easily recognize the potential significance of carcinogens in the air, in water, or even in food, even though we cannot as yet assess their true role or importance. We recognize the fact that the concentration of these carcinogens is increasing. At what point, if at all, do they become significant as factors in the development of human cancers? Certainly, we need the most exact measurements and identification of these carcinogens, not only to help determine their significance but also to measure the need for and efficacy of control programs. Similarly, in our ultimate evaluation of the deleterious effects of radiation, we must have precise basic measurements, even while we are in the stage of speculation and controversy as to the significance of these data.

Currently we are embroiled in a heated public controversy as to the significance of various pesticide, fungicide, or herbicide contaminants of our food and water supplies. A certain female journalist has unduly alarmed the general public with her speculations as to significance, but we know that we did not need to wait for her to recognize the potential hazards attendant upon the use of these chemicals. It is important that we use all possible means to establish the best possible base lines for measurement of the real import of various chemical concentrations and not be forced to rely on the speculative dire predictions of the journalist. Even though some of the base lines that we now attempt to establish may later be found to have little value, we must in our limited knowledge establish as many of these as possible so that we may ultimately select those of greatest value. I include here the baselines that in one way or another may measure the degree to which there are natural or man-made chemical, biological, or physical contaminations of our environment. Concurrently, there rests with us in the realm of medicine a responsibility for developing or improving upon the standards for the measurement of the development, the incidence, and the prevalence of disease. In this latter, you from the environmental field must share with us in the realm of medicine in recognizing that disease and ill health are not simple processes subject to single environmental or metabolic influences, but rather represent a complex interplay of various forces, some destructive, some protective, but all conditioning the ultimate response of the human body.

In our attempt to establish valid environmental factors in the causation of disease, we must not be led into the false assumption that every component of an environmental control program must be evaluated on the basis of a demonstrated relationship to human illness. We must never forget that quite apart from its pathogenic significance an environmental variable may be of import as it affects that vague something that we call the sanitary culture of the community, common decency, or even the aesthetics of human life. The mere fact that I cannot demonstrate a disease relationship for polluted bathing beaches or swimming pools does not alter my reluctance to go swimming in sewage. That I cannot show valid morbidity or mortality statistics as to the significance of excess noise does not make me any less reluctant to live in the cacophony of bedlam. A wise court has long since ruled in a decision on nuisances that demonstration of ill effect is not requisite and that abatement of an environmental nuisance may be required if such renders habitation uncomfortable or interferes with the normal enjoyment of human life. Ideally, we may hope that in as many situations as possible our control measures may be based on exact cause-and-effect measurements, but we must never forget that there are intangibles that contribute to the sanitary culture, the peace of mind, the standard of living of mankind.

As we approach consideration of this problem of environmental measurements, our objectives seem quite clear. On the one hand we must seek ways of measuring the magnitude of a vast array of environmental variables that may conceivably have a bearing upon human health, being still uncertain as to the relative importance or significance of many of these variables. On the other hand, we must attempt to measure the development of human disease so that these findings may be correlated with the results of environmental measurements. In both cases, we must strive for the most precise possible measurements, yet constantly recognizing that statistical correlation does not in itself mean demonstration of a cause-and-effect relationship. Like the pilot of the ancient world beset with mythical maritime dangers in the Straits of Messina, we must attempt to sail the difficult and treacherous course between Scylla and Charybdis, Scylla the multiheaded monster that snatches at and feasts upon fragmentary evidence of no validity and the whirlpool Charybdis that engulfs our imprecise data, churns them through the whirling intricacies of modern electronics, and spews forth the pieces in the form of false conclusions. Wise and skilled is he who safely and successfully sails between Scylla and Charybdis.

SESSION 2: General

Chairman: **Dr. August T. Rossano, Jr.**
Research Professor
Department of Civil Engineering
University of Washington

Dr. Paul B. MacCready, Jr.
President, Meteorology Research, Inc.
Altadena, California

SUMMARY

Principles common to air pollution measurement systems (chiefly meteorological) are outlined, with emphasis on the need for balance between statistical methods and the understanding of physical relationships. Four representative types of measurement-forecast systems are described. Development of a system and application of diverse measurement techniques are exemplified in a hypothetical field study of flow and turbulence regimes at a western hazards site.

DESIGN OF MEASUREMENT SYSTEMS

INTRODUCTION

A measurement system for basic studies in air pollution in the atmosphere or for operational uses can take on one of an infinite variety of forms determined by the program aims, resources in money and people, location, meteorological situation, and the state of the instrumentation art. Since all these subjects cannot be treated meaningfully here, this paper will examine only the basic principles common to most systems and describe a specific project that illustrates some of the principles. Some knowledge of the fundamental principles will permit any given system to be viewed in a useful perspective. The author's background in the subject has derived primarily from many field programs in various weather regimes and terrains, field programs that often had operational aims but necessarily involved some basic physical studies. Therefore in this paper field research systems will be emphasized. The term "system" is here used in a broad sense that means instrumentation, its use, and the handling of the resulting data.

Public Health Service Publication No. 1022, the Proceedings of the National Conference on Air Pollution, is highly recommended as a general reference to this subject, especially the section on "Applying the Measuring and Monitoring Know-How."

SOME GENERALIZATIONS

DATA ACQUISITION

The director of the program using a field measurement system will invariably want more data than any reasonable system is capable of giving him.

ANALYSIS

Nevertheless, usually the data that are obtained cannot be properly assimilated. This is especially true in research projects, for the data treatment is not routine. The data reduction may be easy, but its analysis is not. Virtually every project could benefit materially from more analysis. A reasonable compromise to aim toward on some research programs is to split the funds equally between the instrumentation-field phase and the data reduction-analysis phase.

INSTRUMENTATION SYSTEM

Often the absolute accuracy of measurement is not very important. Making the pertinent measurement at the appropriate place may be more vital, such as being sure the primary wind measurement pertains to the dominant flow. In many studies a network of crude wind recorders will be more useful than one precise unit. In some cases an absolute accuracy of 1°C in temperature is unnecessary, although in other cases an accuracy of 0.01°C in temperature difference is desired.

STATISTICS

The output of most studies or operations is statistical data. The accuracy or usefulness of these data is much improved if the correct sort of physical understanding was involved in deriving the data. This balance between statistical methods and physical understanding is of great importance, and it will be emphasized later in this paper. One main point is that although statistics are involved in typical meterological studies, statistics often constitute only a blunt research tool. Another primary point is that the use of statistics decreases as physical understanding increases. The field of mesometeorology has advanced considerably recently, as have meteorological instrumentation, data processing methods, and turbulence-diffusion relationships; it is now generally possible to interpret the movement and spread of pollution somewhat quantitatively from standard synoptic data. Thus statistical pollution estimates can be refined by physical inputs.

SOME FACTORS IN AN INTEGRATED METEOROLOGICAL SYSTEM

SAMPLING

Meteorological variables and pollution variables cannot be ascertained completely in all three space dimensions and time, but rather samples are taken. The sample is considered to represent the variable over a larger range of time or space If the system sampling design is good, the samples can be truly representative. The sample may be, for example, measurement of wind or pollution at one place at one moment, or the same measurment averaged over a long period; for typical diffusion or pollution studies the latter is more likely to be representative, except for rapidly varying quantities.

AIR MOTION

The movement of pollutants is usually considered to consist of two parts, the mean transport of material and the spread of the material into lower concentrations by means of turbulence. Thus the measurement system must illuminate these two parts. Most commonly the mean flow data would be normal surface wind measurements and tracking of balloon ascents. The turbulence data that is desired actually describes the diffusing power of the air. This diffusing power depends both on the turbulence and on the type of pollution release; a small individual cloud puff is treated by different equations than are used for a continuous point source. The complete relationship between turbulence and diffusion is complex and not adequately understood, but some significant simplifications have been developed recently. These simplifications are based on measuring the turbulence as direction fluctuations of a fast response direction vane in the vertical or horizontal, and processing the analog signal with one or more electronic filtering devices ("sigma" meters) to show the energy over particular broad wavelength bands. For certain cases, such as the diffusion of puffs, only one "sigma" meter is needed, and the rate of cloud spread is simply proportional to the square of the meter reading (at least to the accuracy required in most studies). The turbulence measurements can be made from ground-based equipment or even from aircraft. This turbulence measurement approach is supplanting the older method of inferring the turbulence from measuring the wind and temperature profiles. The turbulence method measures the important parameter directly rather than indirectly, and also it can be used in complex terrain situations where mean profiles are less informative.

In summary, measurement techniques are available now to define the velocity field that carries the pollutants. The measurements show the mean flow of the air and also its diffusing power.

POLLUTANT OR TRACER

In actual pollution work the pollutant is the tracer. Its source is somewhat known. It can be picked up on the ground and sometimes in the air. A tracer can also show what concentrations can be expected at a spot downwind from a particular source, and in addition a tracer is often useful in filling in the picture of just how the material moves from source to destination. The tracer can be a cloud of identifiable particles. Among the many available particulate or gaseous tracers, the most commonly used is zinc cadmium particles of about 2 microns mean diameter, which fall less than 1 meter per hour through the air. The particles are collected by filter or impactor and usually counted by fluorescence under ultraviolet light. Tracing can be done to distances in the hundreds of miles.

An excellent tracer is oil-fog smoke, for by a single visual or photographic observation it can show the entire course of a diffusing mass of air. Although it may be deemed a qualitative tracer technique, in many cases it may actually be more quantitative than particle tracer methods because it can provide many tests simply during one experiment. Particles are measured only at fixed collection points, but the smoke is observed wherever it goes. Smoke from standard generators can sometimes be detected as far as 5 miles from the release point.

A versatile new tracer method is the use of radar to track mylar super-pressure balloons, which float at approximately a constant-density level. Tracking for periods longer than a day has been successful in Los Angeles pollution studies.

ENVIRONMENTAL METEOROLOGICAL DATA

Most pollution meteorological studies hope to provide techniques for forecasting pollution factors from the standard weather data supplied by the U. S. Weather Bureau. During the study the gross environment features are noted, but also the smaller links are examined that connect the flow and diffusion to these environment features — important links such as surface roughness, topography, radiation, and stability.

DATA REDUCTION AND ANALYSIS

The details of the data reduction and analysis depend so much on the specific problem that only a few generalizations are appropriate here. As the automatic weather station concept undergoes continual development, the data acquisition tends to be handled digitally. This puts the data in a convenient form for automatic data reduction. Other data can readily be converted to the digital format, and thus digital data handling can be employed throughout some projects. Automatic data handling is, of course, desirable for routine monitoring programs, and even for research projects it makes analysis easier because more of the pertinent data can be economically provided to the analyst.

The "sigma" meter represents data handling that in many cases is probably best done by analog methods. The "sigma" meter is actually an electronic analog to the running mean of direction variances. Digital techniques for this one task are more expensive than the analog method, and are also less suitable because the digitizing of vane angle is usually done with a resolution (say 1°) that introduces appreciable errors into a "sigma" calculation for weak turbulence conditions.

THE AIM OF A MEASUREMENT SYSTEM

The following list suggests representative types of operational systems:

1. Completely Automatic "Present Picture." This system monitors the three-dimensional field of wind flow and turbulent diffusing power, and can thus present at any moment the picture of the transport and diffusion of a pollutant cloud released at any given point. Since measurements cannot be made everywhere, some empirical extrapolations of data from a few points are actually used.
2. Completely Automatic "Future Picture." This forecasts the "Present Picture," and then derives the flow and diffusion picture. Thus this system must incorporate meteorological forecasting factors such as stability, radiation, topography, and pressure trends. This is the most desired system, but has the basic limitation inherent in any forecasting system in the present state of the forecasting art.
3. Practical Category Type of "Present Picture." Here the most common wind flow and turbulence fields are categorized into a small number of types, and the rules for cloud transport and diffusion are summarized for these types.
4. Practical Category Type of "Future Picture." The flow and turbulence field categories are forecast by theoretical-empirical relationships to available standard meteorological parameters. Then the categories are interpreted, as above, in terms of rules for cloud transport and diffusion. This system is the most practical in the average case. It still has the forecasting limit inherent in any meteorological forecast. It is suitable for compiling statistics on the diffusion climatology of a region.

An example of a pollution research program is given later in this paper. It represents an example of the category types of system outlined in 3 and 4.

DEVELOPMENT OF A MEASUREMENT SYSTEM FOR DIFFUSION MONITORING OR FORECASTING

The key to the development of an operationally useful system lies in basing the system on the simplest possible factors that are dominant in determining the mean flow and turbulence. Thus the success of the system will depend on some physical studies — at least a small field research program is necessary. The system development might follow this course:

1. Decide on the aim of the project or system.
2. Design a tentative operational system, considering likely meteorological factors and taking account of the accuracy obtainable within the economic framework provided.
3. Perform a research program to develop the factors for the operational system.
4. Redesign the final operational system.
5. Keep refining the system as more weather history becomes available.

A SPECIFIC EXAMPLE

The following example of a hazard study illustrates some of the principles inherent in any study or measurement system. The example is a fictitious composite of several real projects, but the results have some general applicability.

The site is considered to be a 25-mile by 25-mile square in the mountain-desert region of the western United States. Toxic materials may be released at two points: one at the top of a 1000-foot ridge, the other halfway up the west face of a 3000-foot ridge to the east of the lower ridge. The problem is to establish the concentrations at which this material might reach the valley east of the high ridge, for the predominant west wind conditions.

The data system consisted of:

1. A portable 10-foot meteorological tower located at the release site or along the course of movement of the tracer, giving mean wind, turbulence, temperature, and temperature gradient.

2. A similar fixed tower on the lower ridge.

3. Four wind stations.

4. A light plane that records altitude, temperature, IR ground "temperature," humidity, turbulence, and rate of climb, with an observer making notes on a magnetic tape recorder and photographing smoke plumes.

5. 40 rotoroid-particle samplers (and 3 filter samplers for backup calibrations) in two lines crosswind to the flow at distances of 3 to 10 miles from the release point.

6. Various generators to dispense fluorescent tracing material (uranine dye for rapid assessment but shorter distances; zinc cadmium sulfide for longer distances), and two oil fog generators.

7. A pibal wind station.

8. Several wind recording systems already at the site (and having provided previous data records).

9. USWB station records at distances of about 25 miles outside the site.

A total of about 30 quantitative tracer tests were made, each one (except for night-time releases) being accompanied by visible smoke releases. In addition, smoke alone was released about 15 times. The smoke releases were of great value in this project because they show so conveniently the mesoscale flow patterns in this complex terrain. In the most convective situations the quantitative tracer samplers on the ground would provide virtually no information; in these situations the visible smoke gave the complete explanation for the lack of counts on the ground.

The tests all took place during one 3-week period, and yet the results can be applied fairly well to other seasons of the year since the flow and turbulence regimes fit rather well into a few identifiable categories. The categories also show a relationship to flows studied on other projects involving waves and turbulence in mountainous terrain.

Figure 1 summarizes the four distinct flow and diffusion categories.

The categories are defined as follows, with the wind speeds referring to velocities at ridge-top levels.

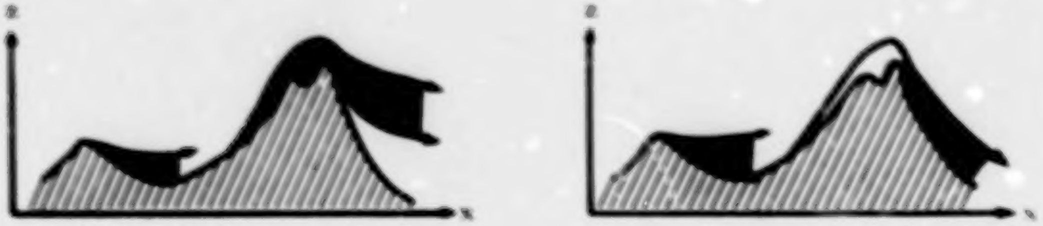
a) **Wind-associated.**
   (1) **with sunny conditions, wind medium to medium-strong.**
   (2) **with cloudy conditions, wind medium.**

b) **Strong wind-associated.**
   (1) **with sunny conditions, wind strong.**
   (2) **with cloudy conditions, wind medium-strong.**

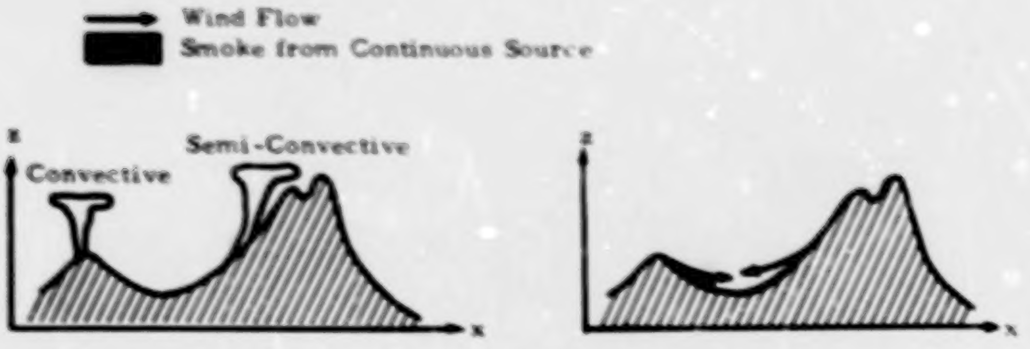
c) **Convective, wind light.**

d) **Semi-convective.**
   (1) **with sunny conditions, medium-light.**
   (2) **with cloudy conditions, wind medium-light.**

e) **Stable, nighttime conditions, low wind speed.**



(a) Wind-Associated (b) Strong Wind-Associated



(c) Convective or (d) Semi-Convective (e) Stable, drainage (night)

Fig. 1 — Schematic Patterns of Flow and Diffusion Categories

The main categories are primarily determined by wind speed, with some stability effect. From USWB records the yearly distribution of these categories can be ascertained to provide a hazard climatology.

For a particular release site and a particular ground pickup spot, the estimated dosage or hazard factor can be derived from the studies by category. Thus the hazard is defined somewhat quantitatively for a given weather situation, the hazard can be predicted to the extent that winds and cloudiness can be forecast, and hazard climatology statistics can be derived with the help of USWB records.

GPO 814-105-2

There are of course more complications than are shown here, such as considerations on wind direction, how to define cloudiness, statistical significance of the estimates, etc. Nevertheless, this study illustrates how practical answers can be derived from a modest program in which the dominant factors are measured.

The program benefited greatly from the versatility of the aircraft, which did stability soundings and turbulence regime plotting, provided a good vantage point for smoke observations, did some of the tracer pickup, and helped to establish gross vertical air motions by vertical velocity measurements and by horizontal plots of potential temperature. Programs in less complex terrain could rely more on ground-based and tower-based equipment.

The categories that came out of this field study could not have been derived from sampler data by statistical analysis, because there were too few tests. It was essentially the physical interpretation of the data that yielded the significant results. The most complete hazard presentation should be given in statistical terms, however. After the sort of preliminary study shown here, in some circumstances it would be warranted to repeat the experiments in meteorological situations representing the greatest hazards, and thus build up statistical data for proper presentation in terms of statistical significance and extreme values.

Dr. Alexander Goetz
Associate Professor of Physics
California Institute of Technology, Pasadena

SUMMARY

Air and water as gaseous and liquid components of the environment are considered essential ingredients for human, animal, and plant life — ingredients that are also acted upon by these live forms. Air and water are evaluated in terms of chemical and physical parameters relating to their occurrence in the natural regenerative and degradative cycle and to their physiological assimiliation. Particulate pollutants and reactive gases are discussed. Emphasis is given to the physical and chemical characteristics of aerosols and their potential role as pollutants of environmental significance.

PARAMETERS

All forms of life exist by continuous interaction with the liquid and gaseous components of their environment, i.e. water and air, for both represent the indispensable vehicles for nutrition and metabolism. One could be tempted to term the function of the vehicles catalytic, but this would be incorrect for, unlike true catalysts, both water and air are gradually degenerated by supporting live forms, regardless of size and complexity, a fact that is the base of all waste problems. The large difference in the physical and chemical constitution of these two vehicles is reflected in the way they serve and are required by specific live units.

Man's average rate of passing the gaseous vehicle is by mass about 3 times that of the liquid (about 7.5 kg or 17 lb of air versus 2.5 kg or 5.5 lb of water in 24 hours), which means that the demand for air on a volume scale is 3000 times greater. Accordingly the type of degeneration caused by the passage through the organism is very different for both vehicles, and it is vastly greater for water than for air.

This environmental degradation by life as such is corrected by natural regeneration, i.e. by processes of neutralization or dispersion, which again differ principally from one another. Water is regenerated by drainage into the oceans, evaporation with subsequent condensation in fogs and clouds, redelivery by rainfall, and subsequent filtration through soil. The atmosphere is regenerated by diffusion, convection, washout through rain, interaction by plant metabolism (CO_2) and — probably on a major scale — by photochemical oxidation (volatilization) of airborne organic constituents.

These natural processes had kept the equilibrium between live matter and its environment on our planet for more than one billion years until this balance was upset by the technological age with a huge variety of artificial energy conversion processes of growing magnitude and the consequence of dense habitation centers, both causing rates of environmental degeneration that rapidly exceed nature's janitorial capacity.

This situation requires corrective action by man, as an answer to the question: "To be or not to be . . ." in the face of the growth rate of these technological endeavors. This action must be effected with minimal economic penalty and thus can succeed only if guided by detailed definition and systematic evaluation of environmental parameters to arrive at standards of general validity.

This evaluation not only consists of the chemical and physical definition of each single pollutant type, as derived from its physiological tolerance limits, but must also critically consider the validity of each such limit in the presence of other pollutant types. Their presence may shift the tolerance for a specific agent substantially in either direction if its effect is synergistically attenuated or intensified.

But even assuming that a system of perfect parameter definitions has been derived from the tolerance pattern, the major problem remains in correlating this system with the indications of the sensor devices available, for it cannot be taken for granted that they always represent quantitatively — truly or even approximately — the reaction of the live organism. Consequently the critical selection of the sensor-types and the judicious correlation of their data represents a major task upon which rests the ultimate success of the effort.

As is to be expected, the parameters pertaining to water and air reflect the principal differences between liquid and gaseous matter and their modes of physiological assimilation.

Contaminants confined to a liquid are much more readily identified and removed at the source than are contaminants dispersed in air.

The lowest tolerance limits for both media are of the same order, ≅ 10^{-8}, which represents in water a mass concentration of 10 μg/liter and is comparable with the molar concentration of 1 pphm in air. Accordingly the same magnitude is required for the maximal sensitivity of the senor devices. These limits are exceeded by several orders for radiological and microbiological pollutants, for the latter because they are potentially self-propagating in many water supply sources — but not in the atmosphere. For these pollutants the sensitivity threshold reaches about one billion times further, for the device must sense, e.g. one *E. coli* cell in 100 milliliters (equivalent to 10^{-8} μg/liter).

The most obvious difference between water and air is the possibility that water can be delivered to the consumer from supply centers where permanent surveillance and corrective action is applied. The water parameters encompass a wide variety of pollutants, present in the dissolved or colloidally suspended state, and their low tolerance limits apply particularly to traces of toxic metals and metalloids, such as arsenic, lead, mercury, manganese, etc.

The highly developed parameter definitions of today for water resources are the result of experience gained during the last 80 years. It is quite interesting to follow the trends of this period: first efforts were to eliminate the visible "dirt," and then to progressively define and control the less and less obvious pollutant traces. This appears to be analogous to our present early state of parameter definition for the air, because it was only about 20 years ago that atmospheric pollutants were generally realized as visible "dirt" in the form of smokes, fumes, and dusts.

Since then we have gone a long way, so that today's pattern of air pollution parameters may be summarized as follows. Principally the parmeters discriminate between contaminants that are temporary and those that are permanent, a distinction that to some degree depends on the rates of atmospheric regeneration as determined by the prevailing micrometeorological conditions. Particulate matter of the larger size classes is considered temporary airborne material because their fallout rates limit their suspension to a few hours. It is difficult to exactly define the lower size limit, because the rate of precipitation depends largely on particle shape and density. In terms of a kinetic (Stokes') diameter the borderline can be placed at 5 to 10 microns, i.e. individual particle masses of 10^{-9} to 10^{-10} gram.

Particulate matter of smaller sizes must be considered as a practically permanent component, because the fallout rate becomes negligible compared to the prevailing velocities of vertical atmospheric convection. This category encompasses the range of aerocolloidal particulates (aerosols). It could be expected that their sizes range downward to those of the individual gas molecules, but present experiences indicate that under normal conditions the lower limit is at about 0.1 micron (in terms of kinetic diameters), since physically definable particulates in natural and polluted air decline rapidly in number below 0.2 micron.[1] This evidence appears to be contradicted by nuclear counts, obtained by adiabatic expansion and subsequent condensation of H_2O in a supersaturated atmosphere, because these nuclear counts indicate in general much higher numbers of smaller particles with sizes far into the millimicron range.[2] The explanation for this fundamental discrepancy may be that this procedure yields "snapshots" of unstable molecular agglomerates with brief statistical existence. One may also postulate that two principal factors cause the absence of stable particles in the two decades above the molecular sizes: the rapidly increasing mobility of smaller particles, which promotes coagulation to larger units and the chemical instability of extremely curved surfaces (Kelvin relation).[3, 4] Both facts would predict a very brief existence for individual particles smaller than 0.1 micron.

Experience has shown that the mass distribution over the total size range is in first approximation constant, which means that the same air volume carries for instance about 1000 times more particulates in the range between 0.2 and 0.3 micron than between 2 and 3 microns.

At the same mass contribution the smaller fraction would present 10 times the particle surface area and mobility of the larger fraction and hence would have about a 100-fold larger chance of interaction with reactive traces in the gas phase surrounding them. This consideration indicates the inadequacy of defining this pollution parameter by mass concentration. As a matter of fact, a truly representative method for assaying the density of aerocolloidal matter appears still to be missing, for it is only partially accomplished by procedures of light-scattering and impaction.

Another form of permanent contaminants, much easier to define, is those in *molecular* dispersion, i.e. gaseous additives to the normal atmospheric constituents. Among these one has to discriminate between two classes: the chemically inert (like CO_2 and CO) and those that can react with other contaminants in the airborne state and form different substances of potentially lesser tolerability. Typical examples of the latter type, though certainly not the only ones, are SO_2 and certain organic hydrocarbons. Their modification by oxidation reactions requires the simultaneous presence of two additional factors: photons in the spectral range of about 320 to 420 millimicrons, as supplied by sun radiation to the lower atmosphere, and ozone (O_3) or oxides of nitrogen NO_X.

According to the well-known pattern[5, 6] NO_X acts in photoactivation as a catalytic activator of the atmospheric oxygen in the presence of gaseous traces that can be subject to oxidation. This leads to the conversion of SO_2 to H_2SO_4, which forms particulates and thereby adds active condensation centers. Similarly NO_X produces a large variety of photochemical oxidation products (oxidants) from a variety of hydrocarbons. Many so-resulting oxidants, particularly those of larger molecular weight ($\geq C_6$), have the tendency to accumulate on existing particulates,[7] thereby causing their growth to many times the original size. This process in turn produces visibility reduction and may cause synergistically intensified irritation, i.e., the typical smog reaction. The particulates present at the time of the reaction thus play a role similar to that of the condensation nuclei for water vapor in the formation of fog.

Detailed studies of such aerocolloidal matter have definitely indicated that these particulate accumulations are not permanent, as evidenced by their gradual shrinkage upon additional irradiation or moderate heat exposure.[8] This suggests that the aerocolloidal pollutants represent intermediate oxidation products, which are gradually converted into more volatile stable end-products such as CO_2, H_2O, and probably NH_3OH. Comparative investigations of aerosols present as "hazes" in unpolluted air (rural, forest, ocean) have indicated this metastability of the particulates to be in a qualitative sense a general property, the main differences from the smog aerosols being the much lower concentrations and the absence of irritant capacity, most likely because of the different nature of the hydrocarbon traces yielded by vegetative life.[9]

For the parameter definition of aerocollodial pollution the general pattern of formation and decay of photoactivated irritants postulates the importance of the initial reaction rate. This pattern should thus depend, under identical conditions of irradiation exposure, on the concentration ratio of the hydrocarbon reactant (HC) to the oxidation catalyst (NO_X), because an excess of the latter should accelerate the decay of the oxidants into their final (neutral) oxidation products. Recent systematic tests[10] appear to support this prediction. The irritation response (which closely parallels the oxidant production) was increased for the same NO_X level with the concentration of HC, while the increase of NO_X beyond 1 ppm *de*creased the response for the same HC level. Similarly, the oxidant formation depended over a wide range on the ratio HC/NO_X, since for the same HC level it decreased when more NO_X was available and vice versa.

This brief resume, although barely outlining the complexity of these reactions, should serve to enumerate and define the parameters pertaining to this type of environmental pollution.

Because the reaction rate in highly diluted systems is slow, obviously time must be available for the irradiation exposure of a given air mass — this means that local conditions of low atmospheric regeneration rates (inversion) are prone to produce such reactions, particularly if sun irradiation is not frequently impaired. Other micrometeorological factors that can affect the photochemical chain reaction rate are, as to be expected, temperature and relative humidity.[11, 12]

Depending on such local conditions, the presence of photochemical activators (NO_X) becomes of primary significance whenever they coexist with hydrocarbon traces. In the enormous variety of this parameter the molecular structure and weight of the organic compounds appears to be of decisive importance; about these factors much too little is known. It is certain, however, that unsaturated compounds of olefinic and/or aromatic structure represent reaction partners of high avidity. It has also been shown that the tendency toward accumulant formation on existing nucleating particulates, i.e. aerosol formation, increases for analogous hydrocarbons (e.g. aliphatic olefins) with their chain length and the asymmetry of their double bonds.[13]

In addition to these reactions, trace pollutant aerosol exists, no doubt, as an independent parameter. The particle number, size-distribution, and type (submicron particulates) characterize the aerosol particles that serve as potential centers of reactant accumulation and reaction-promoting catalysts. Aerosols characterized in this manner have received rather little attention in the past, even though the exhaust of automotive traffic represents a prolific source of particulate matter (lubricants, etc.) for the metropolitan air mass. This is amply evidenced by the fraction precipitated on the surface of the traffic lanes and by the benzene-soluble components of filter deposits.

In a certain sense, the very well-known SO_2 parameter may be considered in the same category. While the irritation records show no statistical coincidence with the SO_2 content,[14] SO_2 is well known as a cause of aerosol formation, because its photochemical oxidation by activating agents produces aerocolloids of high stability and thereby reactive nuclei for the organic accumulants.

The SO_2 parameter requires critical coordination with the coexisting pollutant components in the air mass under consideration. As an isolated pollutant in the absence of reaction partners, SO_2 does not appear to be of major significance.

Finally, another parameter type that appears worth mentioning is sensed by the nuclei-counting devices, which indicate the number of centers for the condensation of water vapor under supersaturation artificially induced. Experience shows that a correlation between *permanent* particles and the frequency of such centers is difficult to establish, if it can be established at all. On the other hand, the nuclei counts seem to indicate the potential reactivity of a system prior to irradiation exposure, i.e. prior to the onset of photoactivation, which gradually declines with the completion of the reaction, i.e. the depletion of reactants. Thus it may be possible to develop from nuclei counting a new parameter that describes the potential activity of the air mass and therefore the likelihood that its chemical composition will be altered by subsequent interreaction of the components.

In summary attention may be drawn once more to the aerocolloidal phase, although it is but an insignificant mass-fraction of the biosphere ($\backsim 10^{-7}$) and thus is not yet accessible to specific chemical analysis, it represents probably a most significant environmental parameter. This parameter is unique, since it does not refer to a defined substance but rather to the "micromorphological" constitution of the air we breathe.

Originating continuously from atmospheric interaction with the planet surface, these submicron particulates modify the statistical pattern of the gas phase by their highly dispersed surface area: they act as centers for H_2O condensation (fog and cloud formation), they also can accumulate more permanent reaction products (hazes), radioactive molecules, etc., and they are important in the photochemical reaction pattern of organic contaminants (smogs). Metropolitan activity changes locally the natural aerocolloid supply, largely by number and chemical constitution. Growth of such nucleating particulates lowers the visibility range by light-scattering or attenuation. The statistical coincidence between irritation and certain aerosol patterns emphasizes the importance of the aerosol parameter. Its detailed definition and application to environment control appears as a primary task of research and development.

REFERENCES

1. A. Goetz and O. Preining, The Aerosol Spectrometer and its Application to Nuclear Condensation Studies, Am. Geophysical Union "Physics of Precipitation", Geophysical Monograph No. 5, 164-182, 1960.
   A. Goetz, The Physics of Aerosols in the Submicron Range, Internat. Symposium on "Inhaled Particles and Vapours", Oxford, England, Pergamon Press 1961, 295-301.

2. C. Junge, Atmospheric Chemistry, Advances in Geophysics, *4*, 1-108 1958. Acad. Press Inc., N. Y.
   H. W. Georgii, Probleme und Stand der Erforschung des atmosphärischen Aerosols, Berichte des Deutschen Wetterdienstes No. 51, 1958 (A comprehensive bibliography of aerosol literature.)

3. G. Zebel, Zur Theorie der Koagulation elektrisch ungeladener Aerosole, Kolloid-Zeitschrift, *156*, 102-107, 1958.
   G. Zebel, Zur Theorie des Verhaltens elektrisch geladener Aerosole, Kolloid-Zeitschrift, *157*, 37-50, 1958.

4. C. Orr, Jr., F. K. Hurd, and W. J. Corbett, Aerosol Size and Relative Humidity, J. Colloid Science, *13*, 472-482, 1958.
   A. Goetz, O. Preining, and H. J. R. Stevenson, Synergistic Properties of Aerosols, Preliminary Report U.S.P.H.S. Grant, Sept. 1958-Dec. 1960.

5. A. J. Haagen-Smit, C. E. Bradley, and M. M. Fox, Ozone Formation in Photochemical Oxidation of Organic Substances, Ind. Engin. Chem., *45*, 2086-2089, *48*, 1884-1887, 1956.

6. P. A. Leighton, Photochemistry of Air Pollution, Academic Press, New York, 1961.

7. A. Goetz, W. Stoeber, T. Kallai, U.S.P.H.S. Grant Progress Report RG 6743, 1961.
   A. Goetz and R. Pueschel, Photochemical Aerosol Formation as a Nucleation Phenomenon, A.C.S. Meeting, New York, September 1963.

8. A. Goetz, O. Preining, and T. Kallai, The Metastability of Natural and Urban Aerosols, Rev. Geofisica Pura e Applicata — Milano, *50*, 67-80, 1961.

9. F. W. Went, Organic Matter in the Atmosphere, and its Possible Relation to Petroleum Formation, Proceedings of Nat. Academy of Sciences, *46*, 212-221, 1960.
   H. W. Georgii, Nitrogen Oxides and Ammonia in the Atmosphere, J. of Geophysical Res., July 1963 (in press).

10. M. W. Korth, A. H. Rose, Jr., and R. C. Stahman, Effects of Hydrocarbon to Oxides of Nitrogen Ratios on Irradiated Auto Exhaust, Part I, Ann. APCA Meeting, Detroit, Mich., June 1963 (Preprint No. 63-19).

11. J. N. Pitts and J. H. Sharp, Effects of Wavelength and Temperature on the Primary Processes of Nitrogen Dioxide, presented at 142nd Meeting, Am. Chem. Soc., Atlantic City, N.J., September 1962.

12. L. A. Ripperton and W. J. Jacumin, Effect of Humidity on Photochemical Oxidant Production, presented at Los Angeles, Calif., A.C.S., Div. of Water and Waste Chemistry, April 1936 and 56th Ann. Meeting APCA, Detroit, Michigan, June 1963.

13. A. Goetz, Methods for Measuring Particle Composition in Photoactivated Aerosols, 56th Ann. Meeting APCA, Detroit, Mich., June 1963 (in press).

14. Technical Progress Report of the Los Angeles County Air Poll. Control District 1962.

Dr. William J. Youden
Consultant, Applied Mathematics Division
National Bureau of Standards
U. S. Department of Commerce, Washington, D. C.

SUMMARY

The effective use of statistical analysis in sampling work is closely tied up with the use of statistically designed sampling plans. Good sampling plans can be designed only by people closely familiar with the features that characterize the area to be sampled. To get the most out of any statistical consultant, we should supply him with results obtained in a small exploratory survey that will give quantitative information regarding the sampling and testing errors. Examples of several basic sampling schemes are presented.

SAMPLING AND STATISTICAL DESIGN

INTRODUCTION

The problem of obtaining a sample that will represent adequately an area or population of interest appears in many forms. Many census and public opinion studies are based on a small fraction of the population. Evaluation of an ore body is made by use of a limited number of borings. The assessment of import duties on wool is based on samples taken from only a small fraction of the number of bales in the shipment. The quality control of manufacturing processes rests on the inspection of samples taken during manufacture. Different as these settings appear, certain considerations are common to all of them.

One of the most basic, and frequently violated, principles is that of the principle of *random* selection of the samples. The price of adopting a *convenient* method of sampling at the sacrifice of a random selection of samples is to vitiate statistical evaluation of the data and to modify the setting of probability limits to the estimates obtained. Whenever systematic sampling is employed on the grounds of convenience, there is no escaping the necessity for first demonstrating that the results check those obtained by random sampling.

SOME GENERAL CONSIDERATIONS REGARDING SAMPLING

By and large there would be no sampling problem if unlimited time and resources were available. But time and resources are usually limited, and the investigator faces the challenge to use these limited resources as effectively as possible.

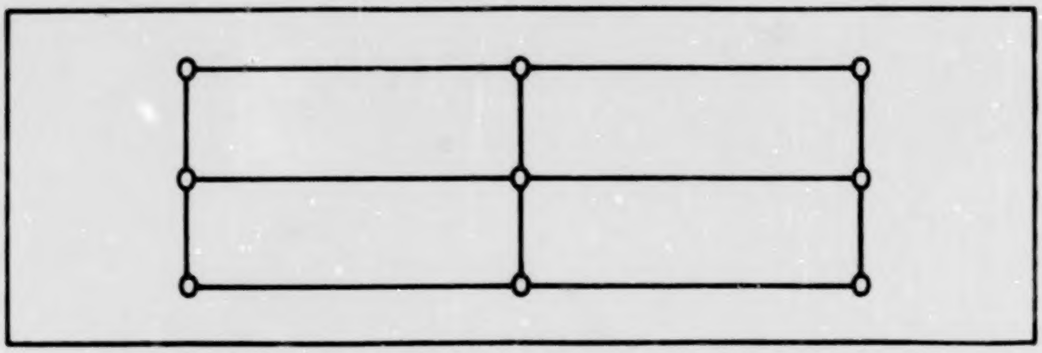
Almost always the element of cost enters into the problem. The inherent value of the samples or the cost of getting them, together with the cost of testing the samples, influence the choice of the sampling scheme. Often provision must be made for keeping tab on the sampling. And always there is the inescapable need first, for a careful prior specification of just what is the region of interest, and second, for a prior decision on the statistical procedures that will be employed. Any sampling investigation that is undertaken with the idea of settling these problems after the sampling is done might as well never be started. In some cases a statistician may be consulted. I wish I could promise you that this was an easy way to obtain a sampling scheme appropriate for the investigation at hand. There is really no easy way, because much depends on the sampling difficulties of the area. Accessibility, uniformity or heterogeneity of the area, and pronounced natural subdivisions within the whole area all exercise a considerable effect on the choice of the sampling plan. The statistician, lacking such detailed information, may make some useful general suggestions but he may miss opportunities to *fit* the sampling scheme to the problem. So really the statistician and investigator should work together, perhaps even explore together the geography or other relevant features of the region. The investigator must pass on the feasibility of suggestions. He should be prepared to give some idea of how accurate he wants his answers to be and he should have available a limited amount of preliminary data to provide information on both sampling and testing errors. All of this is really a minimum for devising a program tailored to fit the immediate problem. This preliminary information often returns many-fold any outlay in time and effort expended before undertaking the final program.

EXAMPLES OF EXPERIMENTAL DESIGN IN SAMPLING

Whole volumes have been written on the topic of sampling. No good would be served in trying to abstract works devoted to the theory and practice of sampling. The available time offers the opportunity to discuss some examples of sampling problems, particularly from the viewpoint of the statistical design of experiments. The design of experiments grew up in the setting of investigations in which the experimenter had many of the important variables virtually under complete control. Thus an experimenter studying the effect of light on the growth of plants could construct an isolated universe wherein he could control the quality, the intensity, and the duration of the light; the temperature; and many other variables that he deemed relevant to his problem. Here the scientist virtually creates the population that he wishes to study, and of course his conclusions are pretty well restricted to this population. Controlled experimentation permits an enormous gain in efficiency compared with investigations of natural populations that necessarily have to be sampled.

There are many problems that cannot be suitably simulated on a laboratory scale, and there is no recourse from the necessity of studying on location, as it were, the phenomenon of interest. The effort is then directed to an examination of the area with a view to ascertaining the actual state of affairs that exists there. Generally speaking this involves a series of point inspections on samples taken at certain points in the region of interest. One sampling technique is to lay down a grid of points or plots, like a checkerboard. The area is divided into rectangles by an equal number of north-south and east-west lines. The spacing between the lines is dictated by the number of samples that can be collected and tested with the assigned resources. Many such programs have been followed. The results permit easy visual representation on a map by drawing contour lines. Often duplicate samples are taken at each point in order to throw light on the adequacy of the sampling technique.

Sometimes little is known about a region, and there is a real necessity to sample the region in such a way that all concerned would accept the sampling as fair. As an example I mention a rather large rectangular slab of concrete on which a large housing unit was to be constructed. The question was raised as to whether or not the slab met specifications, and this could be determined only by boring cores to be tested. Coring is expensive, and even if it were not, no one wants to honeycomb the foundation with holes. The obvious approach was to lay out crisscrossing lines, say three in each direction and take cores, nine in all, at the points of intersection. This is the checkerboard scheme just mentioned.



An alternative approach was suggested that introduced a random element and included parts of the slab closer to the edges. The slab was divided into a 9 by 9 rectangular checkerboard. For ease of presentation here the 81 small rectangular subareas are presented in the form of a Latin Square:

A	B	C	D	E	F	G	H	I
B	C	E	G	D	I	F	A	H
C	D	F	A	H	G	I	E	B
D	H	A	B	F	E	C	I	G
E	G	B	I	C	H	D	F	A
F	I	H	E	B	D	A	G	C
G	F	I	C	A	B	H	D	E
H	E	G	F	I	A	B	C	D
I	A	D	H	G	C	E	B	F

The first nine letters of the alphabet are used to designate the nine subareas in the top row. The same nine letters are used in every row subject to the restriction that when all nine rows have had letters assigned there will be a complete set of letters in every column, i.e., crosswise of the slab.

The Latin Square has been used in agricultural experimentation for 40 years. The idea is that if *n* plant varieties (or fertilizers, or sprays, or other items under test) designated A, B, . . . , G are to be intercompared, it is essential to give every one a fair chance at the available environment. The available area is subdivided into *n* rows and *n* columns. Any treatment, such as C, for example, samples every row and every column, and this puts it on a par with any other treatment. It was found that the accuracy of the comparisons was greatly enhanced by this device of arranging for equality of opportunity for the several treatments. This means that, if in fact all treatments are identical, the average for any one letter should check the average for any other letter. Therefore, from the viewpoint of anyone trying to obtain a representative value for the area, one letter should be as good as any other. Consequently it should suffice to sample just the subareas associated with a particular letter chosen at random. There are a very large number of ways of constructing Latin Squares so that a further element of randomness is also present.

In the actual cement slab problem both interested parties were perfectly willing to abide by the result of samples selected in this manner. This is the real test that the sampling scheme is inherently fair. If more samples are needed, a larger square could be used. Rather than enlarge the square it would be better to choose a second letter, also at random. If the averages for two letters, say D and H, check each other, there is convincing evidence that the sampling is satisfactory. This suggests that the size of the Latin Square should be ½ or ⅓ or other fraction of the total number of samples contemplated. In fact if samples are collected from three letters, the test results on samples from two of the letters may check so well that no tests are needed on samples for the third letter. Incidentally, the samples of any one letter should be identified as to the row and column sampled. Labeling the sample locations makes it possible to compare the average of the samples in the north half of the square with the average for the south half. A similar east-versus-west comparison may be tried.

Twenty-five years ago I was enlisted by a soils man in a study to ascertain the variation in pH of a particular soil type. I devised a sampling scheme that has some interesting features. The approach was simple enough. Two samples were taken ten feet apart. From the midpoint of a line joining these two spots a distance of 100 feet was paced off in an arbitrary, i.e., random direction. Here a second pair of samples was taken. Starting midway between the two pairs a distance of 1000 feet was paced off and another matching set of four samples taken. This set of eight samples was designated a "station." Several such stations were established at intervals of two miles. Table I (Table III in the original publication[2]), shows how the difference between samples depended upon their separation. The interesting thing about this early publication was the noting that, given such preliminary data, a more efficient allocation of the samples could easily be devised.

Table 1. Difference in pH of Duplicate Samples of Culvers Gravelly Silt Loam.

Distance between duplicate samples	0-2 inch layer		2-6 inch layer	
	Av. diff.	Max. diff.	Av. diff.	Max. diff.
10 feet	0.14	0.44	0.11	0.49
100 feet	0.18	0.84	0.20	0.53
1000 feet	0.26	0.69	0.25	0.81
1-3 miles	0.36	1.32	0.28	1.05

The table shows plainly that samples taken close together agreed more closely than samples separated by a considerable distance.

Sometimes the heterogeniety of an area may be quickly demonstrated by a succession of paired samples strung out in line like this:

Pair	1	2	3	4	. . .	n
	. .	. .	. .	. .	. . .	. .

The two samples in each pair of results can be added to give a sum *a* and subtracted to give a difference *d*. If there is no trend the a's and the d's should have the same variance. If there is a trend the a's should vary more among themselves than the d's. We may calculate $\sum d^2$ and $\sum a^2 - (\sum a)^2/n$ with *n* and *n-1* degrees of freedom, respectively. Divide $\sum d^2$ by *n*, call the quotient D. Divide the other quantity by (n-1) and call the quotient A. Then the variance ration, $F = A/D$, may be evaluated by the standard statistical table for *F*. Large values for *F* are evidence of heterogeniety.

A somewhat similiar test may be made to ascertain whether a series of single samples taken in sequence (either along a line or in time) exhibit only random variation. Let the observed results be $x_1, x_2, \ldots, x_n$. Obtain the successive differences $d_1 = x_1 - x_2$; $d_2 = x_2 - x_3; \ldots, d_{n-1} = x_{n-1} - x_n$. Square and sum these differences: $\sum d^2 = D^2$. Calculate $S^2 = \sum (x - \bar{x})^2$. That is, we take the difference between each *x* and the mean *x* and sum the squares of these differences from the mean. In the absence of any trend the theoretical expected value for the ratio D^2/S^2 is exactly 2. If there is a trend along the space or time line, it is natural for adjacent points to be more alike than separated points. Here we would expect the ratio D^2/S^2 to be reduced below 2 because two *successive* samples give only a small chance for the trend to manifest itself. Tables for evaluating this ratio are given by Bennett[1] and in an excerpt from them in Reference 3. If the ratio drops to unity with as few as ten samples, there is evidence at the conventional 5 percent level for a trend.

One widely used sampling device is that of stratification. If the entire region may clearly be subdivided into subregions that are relatively homogeneous within each region, then the allocation of samples can take advantage of this feature with a decided improvement in the formation obtained for a given number of samples. This procedure is well known and will not be discussed here.

I cannot hope to provide even a preliminary list of experimental designs that appear suitable for sampling. There is much to be said for giving the imagination free rein in designing a program instead of limiting oneself to a few standard approaches.

REFERENCES

1. C. A. Bennett, Ind Eng. Chem. *43*, 2063, 1951.
2. W. J. Youden, Contrib. Boyce Thompson Institute *9*, 59, 1937.
3. W. J. Youden, Science *120*, 627, 1954.

Dr. Loren E. Bollinger
Assistant Professor, Department of Aeronautical
and Astronautical Engineering
Ohio State University, Columbus

SUMMARY

Transducers are devices that can be actuated by waves from one or more transmission systems or media and that can supply related waves to one or more other transmission systems or media. Transducers are used to transform one physical phenomenon to another; for example to convert pressure disturbances to related electrical signals.

Major emphasis is placed on transducers used to measure pressure, temperature, and flow rates of various fluids. Transducers used to ascertain chemical composition are discussed also.

The basic principles of operation of the various transducers are given, and the inherent limitations and sources of error are discussed.

TRANSDUCERS

INTRODUCTION

Before any discussion of transducers, their limitations, and their inherent errors, it is well to consider exactly what is meant by "transducer." By one very general definition, transducers are devices that can be actuated by waves from one or more transmission systems or media and that can supply related waves to one or more other transmission systems or media.

On the input side a transducer can convert, for example, a nonelectrical quantity into an electrical signal. A specific example is that of pressure actuating a strain-gage type of transducer that then delivers an electrical output signal, which is some function of the input pressure, to an amplifier or some type of "black box." On the other end of the "black box," an output transducer can change the electrical signal into a nonelectrical quantity, such as the position of a pointer on a meter.

In this paper primary emphasis is on input transducers. Measurements of pressure, temperature, flow rates, and chemical composition are considered. Transducers are needed to sense these primary variables and to change them into corresponding electrical signals, ordinarily, so that appropriate measurement or control can be effected.

Ideally, transducers should respond instantaneously. That is, for a step-function input, the output signal should follow the input variable without distortion of amplitude, frequency, or phase. This concept, of course, is an ideal. No transducer satisfies these requirements over the complete range of feasible input variables. Over a particular, limited range many transducers follow the input variations quite well. Thus, the time element must be considered when transducers are used for measurement.

In some cases there is more concern about the rate of change of the variable than about the magnitude of the change. When the transducer does not follow the input variable exactly, a lag in response results. Lag is the dropping behind or retardation of the output signal in comparison to the input signal. Although the lag of the system may be high, the over-all error could be small. That is, if the error is considered as the deviation of the actual output signal from the ideal output signal, integrated over a long period of time, then the error could be extremely small even though the lag is great. If dynamic error is considered, however, then at any one particular time the error can be very high because of lag. Two types of error should be considered: static error and dynamic error. In a sense, the static error is a deviation of the output signal from the true value of a static variable. Obviously, static error should be held to a minimum. Dynamic error is the deviation of the output signal from its true value when the input signal is varying.

One other major point that should be considered is reproducibility. For repeated measurements of one fixed value of the input variable, reproducibility is a measure of how closely the same output value can be obtained. A high degree of reproducibility is most desirable.

There are many different types of transducers. So that some of these may be examined in detail, the scope of this paper has been limited. The primary transducers considered herein are those utilized to measure pressure, temperature, flow rate, and chemical composition.

PRESSURE

INTRODUCTION

In our world of expanding technology, pressures, both static and stagnation, must be measured in flow fields to satisfy the needs of various industries, scientific and engineering laboratories, and the armed services. Many times the measured pressure can be converted into suitable signal form so that automatic control and regulation systems can be employed advantageously. Before various pressure transducers are discussed, fluids and fluid properties should be clearly understood.

"Fluid" is a comprehensive term that includes two of the three basic categories into which all physical materials are classified generally: solids, liquids, and gases. A fluid can be either a gas or a liquid. Vapors, if considered as a separate classification, are fluids, too. Selection of a specific type of pressure transducer for a particular application depends upon such factors as range, accuracy, frequency response, location of the detector and indicator, reliability, simplicity, availability, fluid temperature, fluid velocity, fluid corrosiveness, adaptability to automatic control, and cost.

The pressure of a fluid, p, is the force per unit area exerted by the fluid on each bounding surface. Within a flowing fluid the pressure may change from location to location because of friction, expansion, contraction, and so forth. At any particular point in a fluid at rest, the pressure acts equally in all directions. Furthermore, the pressure force acts normal or perpendicular to each surface.

In pressure measurement, the force that acts on a known area must be ascertained. In general, pressure is measured by use of two different scales. One scale is absolute in that it is the actual total pressure that acts on a body or surface. When this scale is employed, zero total pressure directly implies an absolutely perfect vacuum. The other scale is relative in the sense that only the pressure above or below the local atmospheric pressure is measured. The barometer is an example of an absolute pressure gage; the conventional Bourdon gage ordinarily uses a relative scale. Analytically, the absolute pressure is the sum of the gage pressure in the vessel and the local value of the atmospheric pressure.

$$p_{\text{absolute}} = p_{\text{atmospheric}} + p_{\text{gage}}$$

MANOMETERS

One of our oldest transducers for the measurement of pressure is the common liquid-column manometer, in many respects the simplest, most direct, and most accurate of all of our pressure-measuring instruments. Unless special manometers are used, the pressure range that can be covered is not great. At very high pressures, manometers become unwieldy; however, they can be used to measure small differential pressures at very high line pressures with great accuracy. Generally, the term manometer is applied universally to a pressure-measuring device that uses a liquid as the measuring medium. There are two principal types of manometers: the U-tube and the well-type.

The simplest manometer consists of a tube of glass or some other transparent material that is bent into the shape of a U. Both legs are filled approximately half-full with a liquid. A modification of the conventional manometer is the inclined tube, in which the manometer tubes are inclined from the vertical for detection of smaller pressure differentials. The fluids normally used for manometers are mercury, oil, or water; different fluids can be used to achieve special effects.

When one looks across the fluid in the manometer, it is obvious that the surface of the liquid is not flat. With some liquids the curvature, the meniscus, is considerable. Liquids that wet the wall of the tubes produce a concave meniscus. Liquids with high surface tension, such as mercury, do not wet the tube wall and produce a convex meniscus. In determinations of pressure difference, a fixed position of the meniscus must be used from one reading to the next. The top of the meniscus is usually read on mercury-filled manometers..

When the U-tube manometer is used, two measurements must be made: the position of the fluid in each leg must be determined. Well-type manometers require the reading of only one leg. In effect, the well-type manometer is a U-tube manometer in which the volume of the second leg is very large. Therefore, the fluid level in the well does not change appreciably as the fluid moves up the vertical column. Corrections in the manometer scale can be made to account for the slight change in elevation of the well-leg.

When manometers are used for pressure measurements, values are commonly given in inches of water or inches of mercury, which are units of pressure. When a column of liquid is subjected to a pressure, the equation describing the equilibrium condition is

$$F_{\text{pressure}} = F_{\text{liquid column}}$$

where F represents the forces. In other words, the pressure force acting on the liquid column is balanced by the force arising from gravity acting on that portion of liquid above the meniscus of the lowest leg. By Newton's second law of motion, it can be shown that the difference in pressure acting on the fluid column is

$$\Delta p = \rho h g / g_c$$

Since g equals g_c numerically, the equation for the balance of forces can be written as

$$\Delta p = \rho h$$

Therefore, for a specific type of manometer fluid, the pressure can be expressed in terms of the height of this particular column of liquid.

To change the pressure range without increasing the height of the manometer tubes, the manometer fluids can be changed. For example, if the pressure range of a water-column manometer must be increased by a factor of 12, the resultant water column could be unreasonably high. If mercury is substituted for water, approximately the same height of manometer tubing can be used because of the change in specific gravity.

Since the density of a liquid is a function of temperature, the temperature must be given when pressure is quoted in terms of liquid head. A reference temperature of 0°C is used commonly for mercury, and 3.9°C (39°F, the temperature at maximum density) for water. Although the pressure may not have been measured at these specified temperatures, the height of the column of liquid at the desired reference temperature can be calculated by use of the ratio of densities at the actual temperature and the reference temperature. Since the pressure is the same at either temperature, the height of the liquid column at the reference temperature is given as

$$h_{ref} = \frac{\rho_{actual}}{\rho_{reference}} h_{actual}$$

It is important to consider that the manometer measures pressure in fundamental units of length and mass. Few other pressure-measuring devices are so basic. By use of various types of oils (silicone, octoil, etc.), the density can be decreased greatly to achieve a greater column height for a given pressure change. When the pressure change is small, very low-density liquids are employed. One of the difficulties with this type of arrangement is that the meniscus of the oil often is difficult to determine.

Capillary effects can cause error when the diameter of the manometer column is too small. In general, the tube diameter should be not less than 10 millimeters for mercury. For water and other fluids that wet the surface of the tubing, the diameter can be somewhat smaller.

With some varieties of well-type manometers a movable well holder is used so that the position of the liquid in the vertical column can be changed. The meniscus of ordinary dater in an manometer is difficult to detect. By the addition of fluorescent dyes such as fluorescein or eosin together with a mild detergent to facilitate surface wetting, the available light is collected and concentrated at the meniscus, which is thereby easier to detect. These additives do not stain the glass tube as do ordinary inks or dyes. Some special manometer fluids, in particular some of the oils, must not be used for measuring pressures of oxygen or oxidizing compounds because of possible chemical reactions. Special devices, such as check valves and traps, can be used with manometers to prevent the liquid from being blown over when excessive pressure is applied accidentally.

PITOT AND PITOT-STATIC TUBES

Generally, almost any combination of mechanical tubes arranged to determine static or stagnation pressures is called a Pitot tube. Strictly, Pitot used the tube to determine stagnation pressures only. Actually there are three basic types of tubes:

1. Pitot tube — a tube, generally cylindrical, pointed directly upstream to measure the stagnation pressure.

2. Static tube — a square-ended tube whose longitudinal axis is perpendicular to the stream lines of the fluid flow, to sense the static pressure.

3. Pitot-static tube — a combination, usually coaxial, of a Pitot tube and a static tube, used to measure stagnation and static pressures at one local region.

In many applications of Pitot-static tubes, the difference between the two pressures is determined directly by a differential-pressure indicator. The difference between the stagnation and static pressure can be measured more accurately with a differential-pressure manometer or similar device than with two independent sensors used to measure each pressure separately. Pitot, static, and Pitot-static tubes are used extensively to measure static and stagnation pressures in wind and water tunnels, on aircraft and marine vessels, and in ducts that carry flowing fluids.

The relationship of static or stagnation pressures, or both, to the other variables makes it possible to calibrate the dials of Pitot-type instruments in terms of the desired variables instead of the pressures that are actually measured. Also, by use of suitable linkage elements in the indicator, functions related to the ratio of static to stagnation pressures can be determined indirectly.

The velocity of a fluid can be obtained by use of a Pitot-static tube from the relation

$$\bar{u} = \sqrt{\frac{2\,\Delta p}{\rho}}$$

where $\bar{u}$ is the average velocity and Δp is the difference between the stagnation and static pressures.

Pitot tubes are fabricated in many different physical configurations, each of which has inherent properties that must be understood, for a particular application. Ordinarily it is assumed that the flow field is one-dimensional, that is, the velocity is a function of one dimension only. When the flow is almost one-dimensional (where the velocity variations in the other two mutually orthogonal directions are small), the error involved by assuming one-dimensional flow is negligible for most applications. Considerable caution must be used, however, when this assumption is made.

If the incompressible, one-dimensional form of Bernoulli's equation for the dynamic pressure is solved, the result is

$$\text{Dynamic Pressure} = q = \tfrac{1}{2}\,\rho V^2 = p^o - p$$

where q is dynamic pressure and V is the average (vector) velocity. When the flow is not one-dimensional, the kinetic energy term q can be written

$$q = \Lambda\,\tfrac{1}{2}\,\rho\bar{V}^2$$

where Λ is a correction factor and $\bar{V}$ is the average velocity over the flow area. The average velocity $\bar{V}$ is a vector quantity composed of the vector sum of the three orthogonal velocity components, u, v, and w in the x, y, and z directions, respectively. Note that q is the kinetic energy of the fluid per unit volume. If something is known about the variation of the velocity as a function of coordinate system, then a theoretical value for Λ may be calculated. If this value can be found, the fluid velocity can be determined more accurately from the measurement of dynamic pressures in multidimensional flow fields.

Suppose, for example, that a fully established laminar flow field exists in a circular pipe. The velocity distribution is parabolic. Setting up and evaluating the various integrals results in a value of 2 for Λ. Thus, the velocities calculated from dynamic pressure measurements would be too large by the factor $\sqrt{2}$.

BOURDON TUBES

The Bourdon tube is usually elliptical in cross-section. Ordinarily it is coiled into a spiral or helix or into a C-shape. In any of the many variations of Bourdon tubes, the free end of the tube moves when pressure is applied internally, and the tubes tend to straighten when the internal pressure is increased. The general tendency is to form a straight, cylindrical tube. When first-order effects are considered, the motion of the free end of the Bourdon tube is directly proportional to the change in internal pressure. Therefore the output function of the device is essentially linear.

The pressure range for Bourdon tubes is from 30 inches of mercury vacuum to 100,000 pounds per square inch pressure. Although many improvements have been made on the basic Bourdon tube, the principle is still the same. Round hollow tubes of suitable material and dimension are flattened to give an elliptical cross-section and then bent into the shape of a C. A tip is sealed onto the free end and the other is connected to a socket that permits connection to the pressure source. With suitable linkage elements, a rack and pinion, and a rotating pointer, the Bourdon tube deflects and causes the pointer to move as pressue is applied.

For a good Bourdon-tube pressure gage, the tube material must be of high quality, with good spring characteristics. Errors can arise from hysteresis in the metal used in the Bourdon tube, from poor material used in the linkage element or in the rack and pinion assembly, and from friction. A diaphragm type of seal can be applied to separate the Bourdon tube from corrosive fluids. Non-corrosive fluids within suitable temperature limitations can be connected directly to the Bourdon tube. If the pressure is pulsating, precautions must be taken to prevent excessive wear or damage to the rack and pinion of the Bourdon-tube gage. Usually some type of pulsation dampener is used to smooth out the pulsations. The inertia of the system limits the frequency response to a low value.

The so-called master or test gages are those that have been fabricated to very high standards of accuracy and can be used for calibration of other gages. It is not uncommon for test gages to be accurate to within 0.25 percent. Temperature variations tend to affect Bourdon-gages too. In some high-precision gages, a bimetal, compensated movement and a hand-calibrated dial are utilized. Accuracy of 0.2 percent or better can be obtained.

RESISTANCE GAGES

Pressures can be measured by transducers with pressure-dependent resistance characteristics. Some variable-resistance pressure transducers have movable contacts; others use continuous-resolution devices. Often the pressure force is converted into an electrical signal by the stretching or compression of a wire (e.g., strain-gage type transducers) or by movement of a sliding contact across a coil of resistance wire, which changes the electrical resistance in the output circuit. Numerous mechanical designs are employed: the resistance element may be a coiled wire, a tapped resistance wire, or a continuous single wire. Carbon strips, an electrolyte, or some liquids, such as mercury, can be employed.

CAPACITANCE GAGES

By use of a movable and fixed metal plate, a variable capacitance gage can be utilized to measure pressure. When a pressure is applied, the capacitance is changed because the distance of separation between the two plates is modified. When a suitable

AC carrier voltage is applied across these plates and fed into an appropriate circuit (usually some form of bridge circuit), an output signal that is a function of the pressure can be obtained. Capacitance gages can yield fairly good transient response. This type of transducer does suffer from temperature effects unless special low-expansion metals, such as invar, are employed. If the capacitance probe is water-cooled, the effects from temperature changes can be minimized.

The air gap between the movable plate and the fixed plate is small, for example, 0.003 inch. The displacement during application of pressure is approximately 1/10 that value. A dielectric other than air — for example, mica — may be substituted between the plates. For reproducibility of data, the two plates must be kept parallel. Some special capacitance probes have natural frequencies as high as 500 kc, but their use is limited.

PIEZOELECTRIC GAGES

When a force is applied to certain types of crystals along specific planes of stress, the crystal produces an electrical charge. When the crystal is appropriately coupled to the pressure system, an electrical output signal can be obtained simply by allowing the deformation force (pressure) to act on the crystal. Electrical contacts are made to the crystal, and the delivered charge, which is a function of the pressure, is measured. Typical crystals are quartz, tourmaline, ammonium dihydrogen phosphate, barium titanite, and Rochelle salts. Quartz crystals, either natural or synthetic, are often used because they allow very low electrical leakage and permit the measurement of slowly varying pressures.

The output signal must be fed into an extremely high-impedence amplifier to decouple the crystal effectively. The charge produced per unit pressure is low. Usually the input resistance to the amplifier ranges high in megohms. Some type of electrometer circuit is employed ordinarily, in which case the input resistance is usually higher than 10^9 ohms. The quartz pressure transducers are very useful in measuring transient pressure waves that have very fast rise times. With special care it is possible to measure accurately the transient pressure of a shock wave having a rise time in the microsecond region. About the only practical method of displaying the output from such a device is to use an ocilloscope with high-frequency response. The trace can be photographed with a camera.

TEMPERATURE MEASUREMENTS

INTRODUCTION

Temperature is an intensive and not an extensive quantity. No unit temperature interval can be applied successively to measure any other temperature interval, as can be done in the measurement of such quantities as length or mass. The size of the degree on one part of the scale, no matter how well defined, can bear no relation to the size of the degree on any other part of the scale.

Temperature scales based on different thermometric substances or thermometric properties differ fundamentally. The difference between two scales that differ only in function chosen is superficial because the conversion from one scale to another is merely a matter of calculation. If the same basic fixed points are used, the scales will necessarily agree at these points but not at others. For example, two scales, both based on the apparent expansion of mercury in glass, will differ unless the type of glass used is identical.

LIQUID-IN-GLASS THERMOMETERS

One of the simplest temperature-measuring devices is the common liquid-in-glass thermometer, in which mercury is often used. The basic principle is the use of the volumetric expansion of mercury as a function of temperature as a means of indicating temperature. The glass thermometer or glass tube has a bulb formed by a glass envelope, which contains the mercury deposited in a metal or glass well at the bottom. When heat is applied to the thermometer, it is transferred through the wall into the mercury. As the mercury expands the column rises in the capillary tube.

Temperatures can be measured by calibration of the position of the mercury in the glass tubing as a function of temperature. The expansion and contraction of the glass envelope must be considered when the calibration marks are etched on the glass. Some thermometers are made for partial immersion, usually 3 inches, or for total immersion. The scales on the tubes ordinarily are calibrated for one or the other condition. By shaping of the glass stem, magnification can be incorporated for easier determination of the position of the mercury meniscus.

The space above the mercury column generally is filled with pure nitrogen under pressure. The gas above the liquid mercury tends to minimize breaking of the mercury thread when the thermometer is handled roughly. Also, the increased gas pressure above the mercury raises its boiling point.

Pointing of quality glass thermometers consists of placing file marks on the stem. A five-point thermometer is calibrated at five fixed points.

It is essential that the bore of the glass thermometer is uniform and that the mercury is pure. Readability is improved by use of color contrasts such as black, yellow, ruby glass, white glass, etc. An enormous ratio of bulb volume to bore volume gives good precision but not necessarily good accuracy. The finest test-grade glass thermometers can be read to within approximately 0.02° with engraved graduations of 0.1°.

Since glass ages regardless of precautions, the stability of a themometer is affected somewhat with age. Elasticity is the primary property of concern. Exposure of the bulb to much higher or lower temperatures than those for which the thermometer was designed can upset the aging process of the glass. Manufacturers allow for the aging process in design of the thermometer.

The response time of the liquid-in-glass thermometer is one of the longest among ordinarily used temperature-measuring devices. Calibration can be affected if the bulb volume changes with time. The change is generally less than 0.1°C for a good grade of glass if it has not been used above 150°C. Hysteresis can be noted when thermometers are heated to or above 150°C and then cooled and checked. The thermometer will read low because the volume has increased. Many times the thermometer will return to its original calibration in a few days. Nitrogen at 1 atmosphere pressure above the mercury is used for measurement of temperatures up to 300°C; 20 atmospheres of nitrogen are used for temperature measurements as high as 550°C. The softening point of glass must be considered.

For differential temperature measurements, the Beckman thermometer can be employed. The expansion chamber of an ordinary thermometer is enlarged so that mercury can be poured into it from the main reservoir. The range is usually from —35° to +300°C; a differential range of 5°C with readability down to 1/100, or over 1/1000°C, can be obtained. The scale lengths are available up to approximately 30 centimeters.

In many measurements the temperature difference is more important than the absolute temperature.

BIMETALLIC THERMOMETERS

By coupling two metals that have different rates of expansion with temperature, the temperature can be measured by observing the deflection of the free end of the combined strip. The bimetallic thermometer is a rugged and simple device for the indication of temperature. The accuracy is not high.

The bimetallic strip can be made in the form of a straight cantilever beam; a change in temperature causes the free end to deflect, and this movement can be calibrated. Generally the deflection is nearly linear with temperature. In other transducers, the bimetallic strip is wound in the form of a helix; one end is fastened permanently to the case while the other is attached to a pointer on a dial. Commercial bimetallic thermometers generally cover the range from —40° to +425°C.

RADIATION PYROMETRY

Radiation and absorption is a universal process of heat transfer. Radiant energy travels from a source or a radiator until the energy is absorbed by the medium in which it is traveling or is intercepted by an object. Energy may be transferred from one body to another by the process of radiation and absorption even though there is no material in the space between the bodies. Upon interception, the energy is partly reflected, partly absorbed, and partly transmitted.

All bodies emit radiant energy at a rate that increases with temperature and is independent of the neighboring bodies. A "black body" is a body that absorbs all radiation incident upon it and reflects or transmits none. A black body is an ideal radiator. It emits, at any specified temperature, in each part of the spectrum the maximum energy obtainable per unit time from any radiator as a result of temperature alone. Often it is convenient and desirable to measure the temperature of the surface of a body by means of the neutral radiant energy emitted from it. One need not make any connection to the body or be in close physical proximity to it. Measurements can be made on moving bodies, corrosive liquids, and distant objects at high temperatures. Radiant flux is the rate of flow of energy from a radiator.

Let P = radiant flux

U = radiant energy

Then $P = \dfrac{dU}{dt}$ (ergs/sec or watts)

Let R = Radiance of a source $= \dfrac{dp_e}{dA}$ (ergs/(sec • cm²))

e = emitted from source

The radiance of an actual source is related to that of a black body by the total emittance ε.

$$\varepsilon = \frac{R}{R_b}$$

where b refers to a black body and ε is also called the emissivity of the body. It is a measure of the deviation from a perfect radiator. One of the pertinent radiation characteristics is stated as Kirchhoff's Law.

The emittance ε of a non-black body is equal to the total absorbtance α for radiation from a black body at the same temperature. $\varepsilon = \alpha$ and $\varepsilon\lambda = \alpha\lambda$ where λ is a particular wavelength. No material is a true black body. Some solid bodies can be converted into artificial black bodies by drilling them with a small hole or a wedge. By use of radiation from the hole or the wedge, black-body radiation can be approached. By the Stefan-Boltzmann total radiation law, $R_b = \sigma T^4$, temperature can be determined by measuring the total radiation from a black body or a gray body, which is one that deviates by a known and constant amount for any wavelength from a true black body.

The indication or deflection of an instrument depends on the surrounding temperature T_0 and the temperature of the body T. Thus, $D = C_1 \ (T^4 - T_0^4)$ where C_1 depends on the instrument used and the physical arrangement of the heated body.

Total radiation transducers are useful for measurements at low temperatures. To provide sufficient output, the radiation detector can be formed of a collection of thermocouples, connected in an additive arrangement known as a thermopile. At high temperatures, say above 600°C, where the object glows visibly, an optical pyrometer can be used.

Often the disappearing-filament type of optical pyrometer yields good results. The radiation from a black body or known gray body is focused with an optical system. The observer sees the radiator and a heated wire filament, which is located within the pyrometer. The filament is a tungsten wire, heated with a battery; a series variable resistor is included so that continuous changes can be made in the temperature of the filament.

By sighting on the black body hole, one sees both the hole and the heated filament. The current through the filament is adjusted until its color temperature matches that of the black-body radiator. Then, when the image of the filament is moved slightly in the pyrometer across the hole, the filament tends to disappear when the color temperatures have been matched. The temperature reading is obtained by examining the scale connected to the variable resistor that controls the current through the filament wire. In turn, this scale and filament color temperature are calibrated against a standard tungsten lamp whose radiation characateristics are well known.

Accuracy of measurement depends on how well the observer is trained and on the quality of calibration of the pyrometer against a standard tungsten lamp source. The National Bureau of Standards calibrates optical pyrometers over the temperature range from 800° to 4200°C. The uncertainty of calibration varies from 3° at the gold point to 40° at 4200°C. Also, the standard tungsten strip lamps, which are used as sources of known brightness temperature over the range from 800° to 2300°C, can be calibrated.

Another type of pyrometer employs two selected wavelengths to obtain temperature measurements. The optical disappearing-filament-type pyrometer uses a single filter whose wavelength usually is centered at 6500 angstroms. With the ratio pyrometer, two wavelengths are chosen and a ratio is formed from the two output signals. This ratio, which is a unique function of temperature over a wide range, is calibrated as a function of temperature.

Some experiments are being conducted with a photoelectric pyrometer, which replaces the human-eye detection system with a photomultiplier tube to make brightness matches between the subject and the pyrometer lamp. This instrument should eliminate the variability in calibration caused by the observer's lack of precision.

RESISTANCE THERMOMETERS

Since the resistance of most metals is temperature dependent, a thermometer can be made by winding a resistor with a selected metal. Then, by accurate measurement of the resistance, the temperature can be determined. The resistance bulb can be used to measure the absolute temperature because the resistance of the wire in the coil depends directly on temperature. Resistance thermometers can have high sensitivity; that is, the change of resistance per degree is appreciable. There is a maximum temperature limit, however, above which the resistance bulb cannot be used.

The wire material must not undergo any phase changes during the temperature excursion, or its characteristics will be changed and the calibration altered. Normally resistance thermometers are relatively large compared to thermocouples or thermistors. Resistance thermometers are made from a variety of materials, often nickel with platinum and copper. The temperature-resistance curve of nickel is non-linear. The shape of the thermometer varies greatly depending upon application. For accurate readings, all portions of the resistance thermometer must be at the same temperature.

The resistance of the thermometer is measured frequently with one of a number of bridge circuits. In some cases the resistance bulb can be made an integral part of a bridge circuit that incorporates the slide wire of a chart or indicating-type recorder. Platinum is often selected because of its excellent reproducibility from —260° to 1100°C. Nickel is limited generally to use at temperatures below 300°C. Below —260°C, platinum becomes a super conductor.

The wire in the resistance thermometer generally is doubled upon itself to preclude inductive effects. The wire must be free of supports to minimize heat conduction losses. Protection tubes, either metal or ceramic, can be used if required. Metal tubes often are filled with a dry gas, such as air or nitrogen, at approximately 0.5 atmosphere pressure at room temperature. Some pressure should be maintained in the tube to increase the rate of heat transfer and thus improve the response time. Accuracy of approximately 0.001°C can be obtained without extreme care. Some care must be taken with the leads coming from the resistance element, since they can affect the accuracy because of the effect of the temperature gradient on their resistance.

THERMISTORS

One of the newer transducers for temperature measurements is the thermistor, a resistor that is extremely sensitive to temperature. Thermistors show a high negative coefficient of resistance as a function of temperature. It is not uncommon to find a semiconductor material that changes its resistance by a ratio of 10^7: 1 over the temperature range from —100° to +450°C. The use of thermistors at temperatures above 450°C and below —180° is uncommon.

The normal resistance value of thermistors at ambient temperatures varies widely; sometimes it is only a few ohms, sometimes a few megohms. Many semiconductive materials can be used to fabricate thermistors. These include some of the metal oxides and a number of mixtures. Thermistors can be made in extremely small sizes and in odd shapes. The method of measuring the resistance of a thermistor must be selected carefully because current from the resistance-measuring device can change the junction temperature and thereby give a false value because of the great sensitivity of the junction resistance to temperature.

The sensitivity of resistance to temperature varies widely, but values ranging from

1 to 5 percent per degree centigrade near ambient temperature are not uncommon. These values often increase at low temperatures, and decrease at high temperatures. Thermistors are useful in temperature control devices because of their high sensitivity.

THERMOCOUPLES

In 1821 Seebeck discovered than an electric current will flow in a closed circuit when two dissimilar metals are used and the temperature of one junction is hotter than that of the other junction. In 1834 Peltier discovered that when a current flows in one direction across the junction of two dissimilar metals, heat is absorbed and the junction is cooled. If the direction of current is reversed, the junction is heated instead of being cooled. This phenomenon is reversible.

The heat developed in the junction is a function of the first power of current rather than the conventional I^2R — Joule heating, which is irreversible. The Peltier heat depends only on the pair of metals chosen and is independent of the form and dimensions, whereas Joule heating is a function of form and dimensions. The amount of current that flows as a result of the junction of dissimilar metals depends primarily on the temperature difference, the choice of metals, and other factors, including the total resistance of the circuit.

If an open circuit is used, the potential difference that will exist between the terminals will depend on the temperatures at both ends of the couple, but not on the shape or the dimension of the conductors. When two metals are placed in contact, electrons diffuse across the boundary continuously until an electric field is established whose force opposes the transfer of more electrons, thereby establishing an equilibrium condition. This output voltage is a function of the temperature difference and the absolute temperature of the cold junction. Since the voltage of a single junction cannot be measured alone without introducing additional junctions, at least two junctions exist in any practical thermocouple circuit.

Thomson deduced that the Peltier effect was not the only reversible heat effect, but that there is a reversible effect within the conductor itself when there is a temperature gradient and a current. Later he proved it experimentally.

In summary, the following effects exist:

1. *Seebeck Effect* — Electric current flows in a closed circuit if two dissimilar conductors are used when the temperature of one junction is higher than that of the other junction. The Seebeck effect is the sum of the Peltier and Thomson effects.
2. *Peltier Effect* — Electric current flowing in one direction across the junction of two dissimilar metals causes heating or cooling. The amount of heating or cooling is directly proportional to the quantity of current. When the direction of electric current flow is reversed, the heating and cooling effects are reversed. Peltier heat depends on the type of metal. The amount of heat is independent of material form and dimension.
3. *Thomson Effect* — The Peltier effect is not the only reversible heat effect. Thomson concluded that there must be a reversible effect within the conductor itself if there is a temperature gradient along the metal conductor. The temperature rises for cadmium, silver, and zinc for a particular direction of electric current. For the same direction, the temperature drops for iron and nickel. The temperature change is zero for lead.

In any simple thermocouple circuit consisting of two junctions and two wires, a minimum of four voltages exist; two Peltier emf's appear at each junction and two

Thomson emf's appear along each wire. Ordinarily the Peltier voltage is less than 0.1 volt. At low temperatures the voltage output is extremely low, in the microvolt region. Originally, most thermocouple tables were prepared with lead as the reference metal because lead has a zero Thomson coefficient. Nearly all modern tables use platinum as the reference because it melts at a much higher temperature.

In conventional thermocouple usage, one of the thermocouple junctions is maintained at a fixed reference temperature, which ordinarily is a well-prepared ice bath. One cannot prepare an ice bath by placing ice cubes in water and expect to have a good, stable, known reference temperature. Pure water must be used, with the proper amount of air saturation. The ice must be crushed well and be in good contact with water, because the definition of the ice point depends upon equilibrium between ice and water.

The temperature of the cold junction is known as the reference temperature. For special applications liquid hydrogen or liquid nitrogen may be used instead of ice to provide the reference temperature, and at high temperatures the boiling point of liquid sulphur may be used. Under these conditions data are difficult to interpret because separate calibration curves must be employed, since the thermoelectric power changes with absolute temperature. Thermoelectric power, which is a misnomer, is the rate of change of voltage with respect to temperature, that is, dE/dT.

If the temperature of the hot junction is raised sufficiently high without the metal melting, ordinarily a neutral temperature can be reached. At this point, the voltage no longer rises with increasing temperature; the slope of the voltage-temperature curve iz zero. A further increase in temperature causes a decrease in output voltage; finally an inversion temperature is reached, at which the output voltage is zero. Still higher temperatures cause the output voltage to reverse polarity. Ordinarily thermocouples are never used even as far as the neutral temperature.

The number and type of themocouple materials are manifold. Conventional types are listed in many handbooks; these data are readily available. For special high-temperature applications, however, one must use some peculiar metals. Platinum and platinum-rhodium alloys can be used at temperatures up to approximately 1800°C with care. Above that temperature pure tungsten, tungsten alloys, rhenium, and other refractory materials, including tantalum, molybdenum, and iridium can be employed.

Reproducibility is not good, however, and the thermocouples are difficult to calibrate. Many of these metals are extremely sensitive to oxidation at high temperatures.

Within ordinary temperature ranges, say several hundred degrees above and fifty degrees below the freezing point of water, temperatures can be measured very accurately with thermocouples. Adequate low-level measuring devices, of course, must be employed. By use of extremely small-diameter wire to form the junctions, very high-speed transient responses are achieved, sometimes in the microsecond range.

Three basic laws apply to thermoelectric circuits.

1. Law of homogeneous circuits. An electric current cannot be maintained in a circuit composed of a single homogeneous metal, regardless of the cross-sectional area, by the application of heat alone.

2. Law of intermediate metals. If a number of different thermocouple junctions exist in a circuit and the entire circuit is maintained at one temperature, the algebraic sum of the thermal voltages will be zero.

3. Law of intermediate temperatures. The electromotive force generated by a

thermocouple having its junctions at dissimilar temperatures T_1 and T_2 is equal to the algebraic sum of the electromotive forces developed by the thermocouple with junctions at T_1 and T_3 and the same thermocouple with junctions at T_3 and T_2.

FLOW

INTRODUCTION

Almost anything that provides resistance to flow can be made into a flowmeter. The pressure drop across the flow resistance can be calibrated in terms of flow rate. For general usage the calibration must be repeatable, the transducer must be sufficiently sensitive, the flowmeter must withstand the action of corrosive fluids, and the flowmeter must give adequate frequency response.

Most flow measurements are made by inferential techniques; that is, pressures or positions are measured and the flow rates are inferred. Although almost any kind of restriction can be used for a flowmeter, it is desirable to use one for which published coefficient data are available so that the pressure drop can be predicted for a given flow rate without calibration of the flowmeter.

HEAD FLOWMETERS

The basic principle of head flowmeters is the conversion of energy from one form to another by the primary element. The conversion from kinetic energy to potential energy is made in flowmeters for liquids; the liquid is essentially incompressible. An average flow measurment is obtained with the common head meter because of the difficulty of making point measurements. When the flow rate of a liquid is measured, the pressure difference across the head meter is a function of the velocity, density, and viscosity of the flowing stream. It is generally assumed that the fluids are homogeneous; otherwise most of the existing coefficient data would not be applicable. Nonhomogeneous fluids give considerable difficulty in interpretation. When the fluid is a compressible gas, the internal energy of compression must be considered in the energy conversion.

A number of primary elements can be employed to convert some of the energy from kinetic to potential. One of the most common is the thin-plate, square-edged orifice. It is easy to install, to inspect, and to replace if damaged or to substitute if a change in flow rate is required for a particular value of differential pressure. The orifice can be reproduced readily, although appreciable care must be taken in fabrication of the plate. Beyond certain minimum sizes, the characteristic coefficient depends primarily on the ratio of the diameter of the orifice to the diameter of the pipe. Tables of coefficient data are readily available in the literature.

In general, there are four basic types of orifice plates, each designed for particular applications. The most common is the concentric type, in which the bore is located concentrically with the inside of the pipe. This type is used often when the fluids are clean and the gases contain little or no liquid. If the gas carries some liquid or solid material, a vent or drain hole can be provided.

In some applications the liquid contains a large quantity of undissolved gases or the gas to be measured contains a considerable number of condensible components that are carried along in the pipe. For these conditions an eccentric orifice plate can be utilized. The hole is located tangent to one wall of the flowmeter tube. The eccentric orifice is similar to the concentric orifice except that the hole is located off-center and the outer portion is tangent to the pipe wall. Thus the flowmeter is fully vented and fully drained. This type of a flowmeter is less accurate than the concentric type, as evidenced by the poorer coefficient data in the literature. If the flowmeter can be calibrated with the fluid to be used, the eccentric orifice is as accurate as the other types of flowmeters.

The segmented orifice plate is useful for liquids that carry solids in suspension. The segmented orifice plate covers the upper cross-section of the flowmeter pipe. The lower section is left completely free so that the solids will not accumulate on the upstream side of the orifice plate. The chord section is fabricated with a sharp edge; the rear portion has a radius of curvature whose arc is 98 percent of the nominal pipe radius, so that the curved surface will not be located below the wall surface of the pipe.

The quadrant type of orifice plate is useful in special cases. With sharp-edged orifices the flow coefficient increases as the turbulence decreases. In many situations the flow becomes so highly turbulent that this coefficient change is of no significance. When the fluid viscosity is above five centipoise, however, and the quantity to be transferred through the flowmeter is relatively low, the concentric orifice does not operate satisfactorily because of the large change in flow coefficient with flow rate. In this case the quadrant type of orifice is useful. This primary element shows little or no change in the coefficient for low turbulence conditions. With this plate the curvature on the approach side is the quadrant of a circle and the radius of curvature depends on the throat ratio.

Manufacture of quadrant plates is difficult because of the curvature requirement. Since the nature of the surface greatly influences the flow coefficient, it is usually necessary that a field calibration be made. The quadrant plate does exhibit excellent reproducibility. For flowmeter pipe diameters of less than 2 inches, the published coefficient data become somewhat questionable. Therefore, it is highly recommended that flowmeters of smaller pipe diameters be calibrated.

In a venturi tube, the fluid-carrying pipe contracts and expands gradually to form a smooth convergent-divergent nozzle. According to the basic flow equations, the fluid accelerates as it passes through the venturi. Although venturi installations are more expensive and more difficult to fabricate than orifice plates, they are useful when extremely large quantities of fluids are to be measured. The pressure recovery with the venturi tube is excellent, and the inner surface of the tube is smooth. The recovery section generally is designed with a 1-to-10 taper; that is, the diameter increases 1 inch for each 10 inches of length, giving a 20-to-1 slope on each side. Steeper or less steep slopes result in reduced pressure recovery. With small venturi tubes, the Reynolds number and viscosity affect the measurements. The surface finish and irregularities become more important as the size decreases.

A flow nozzle is essentially a type of venturi tube without the recovery section. The converging section is generally short.

AREA FLOWMETERS

In the area flowmeter a float is positioned in a variable area section. The fluid enters the bottom portion of a tube and passes upward through the metering section around a float positioned concentrically within the variable area tube. The fluid then exits at the top of the tube. The metering tube is usually glass, with sides tapered uniformly so that the cross section at the top is greater than that at the bottom. The float inside is guided so that it moves up and down concentrically within the tube as the flow rate changes. This type of flowmeter must be installed in a vertical plane.

When the flow rate is steady, the float assumes a fixed position within the tube and its position can be calibrated in terms of the fluid flow rate. The float is in static equilibrium for a constant flow rate because of the equality of the gravity and bouyancy forces. Some flowmeters of this type incorporate a heavy and a light float and thus constitute a two-range flowmeter within one tapered-tube housing.

Variable area flowmeters are produced in many styles. Some are armored for use at high pressures; others include an electrical readout system so that the output signal can be fed directly into a recorder, an indicator, or a computer.

WEIRS, FLUMES, AND NOZZLES

Head-area flowmeters often are used in open channels to measure liquid flow rates. These open-channel meters are commonly used in electrical generation stations, water works, sewage disposal facilities, and water irrigation. One of the most common types of weirs incorporates a rectangular notch; in other weirs the notches are V-shaped or trapazoid. In the rectangular-notch weir, the velocity is proportional to the depth at the weir.

For the calibration of weirs approximate discharge coefficients are available in the literature. The flow rate through the weir depends on the height of the fluid flow in the rectangular opening raised to the 3/2 power. The exponent on the elevation term varies with other shapes of weirs.

Flumes are used in open streams where the flow rates are much greater than those that can be measured with a weir, up to 70 million gallons of water per day, for example. The flume restricts the stream and then expands it again in a definite fashion. The head is measured at a single point about one third of the distance downstream from the inlet of the flume in the entrance section. Either test data or an empirical formula must be used to obtain the flow rate.

An open nozzle can be used to measure the flow rate of sewage, sludge, and industrial waste in pipes and channels that are partially filled. The unique cross-sectional shape of the nozzle produces a nearly linear relationship between head and flow.

ELECTROMAGNETIC FLOWMETERS

The concept on which the electromagnetic flowmeter is based is that of the electric generator. When a conductor moves in a magnetic field, the voltage generated in the conductor is proportional to the strength of the magnetic field and to the velocity at which the conductor moves. For the measurement of fluid flow rates by an electromagnetic flowmeter, the fluid must be a conductor having a reasonable electrical conductivity. A uniform magnetic field is produced either by a permanent magnet or by an electromagnet located outside the pipe. The generated voltage is measured by a pair of insulated electrodes located in opposite sides of the pipe on an axis perpendicular to the magnetic field.

When the magnetic field is uniform, the voltage developed is proportional to the velocity of the fluid flow. There are no obstructions in the pipe. The electromagnetic flowmeter is useful in measuring the flow rates of liquefied metals, particularly those used in the nuclear industries. For some water-based liquids, use of an AC magnetic field is desirable, because of polarization difficulties at the electrodes.

Difficulties are encountered when the conductivity of the fluid changes. The response time is excellent. Since it depends on the frequency of the magnetic field, a high-frequency field must be used when extremely high response is required. Electromagnetic flowmeters are suitable for measuring the flow rates of corrosive fluids and slurries. Results are relatively unaffected by viscosity, density, and turbulence.

DISPLACEMENT FLOWMETERS

The liquid displacement flowmeter calibration method is one of the primary standards for gas flows. This technique consists of displacing a known or measurable volume of a liquid with the gas at a known pressure and temperature. The flow rate of the gas must be constant with this technique unless only the integrated value of the total volume of gas is needed. The volume of the displaced liquid can be calculated by weight or by volume over a known interval of time to yield an average volume flow rate of a gas at the conditions of test.

In some displacement flowmeters the level of liquid remains relatively constant; in the positive displacement type, some portion of the gas stream energy is required to move the liquid. Generally, this type of flowmeter measures relatively low flow rates with a high degree of accuracy. The wet test meter is a displacement flowmeter which uses liquid displacement.

TURBINE FLOWMETERS

With turbine flowmeters, a free-running turbine is mounted in the flow stream. By use of special shapes for the turbine to reduce inertia, friction, and viscosity effects, the flow rate is determined by measuring the rate of revolutions of the turbine. Ordinarily, a magnetic-type pickup is employed, and the output pulses are fed to a pulse-counter system in which the pulse rate is determined electronically. This rate is a function of the mass flow rate of the fluid flowing through the meter.

The advantages of the turbine flowmeter are that the pressure drop across the flowmeter section is small, the line element is very compact, and the flowmeter can operate at temperatures from below zero to above 500°C. They range in size from 1/8 to more than 8 inches diameter. Response times of the turbine-type flowmeters are extremely rapid, and accuracy of 0.1 percent can be achieved under special conditions. Flowmeter units must be interchanged when the change in ranges is more than about 10 to 1.

CHEMICAL COMPOSITION

INTRODUCTION

In many applications the chemical composition of materials must be determined. Since many techniques are available for determining the composition of fluids or solids, the choice depends upon such factors as the material, the available instruments, the cost, and the accuracy required. There is no systematic method by which the best possible technique or transducer can be selected to determine chemical composition. One can use transducers based on electromagnetic radiation (including X-rays), chemical affinity or reactivity, electric or magnetic fields, thermal or mechanical energy, and other principles.

SPECTROSCOPY

Spectroscopy is the measurement of the position of the wave length of interest within the spectrum and its relative or absolute intensity. Both emission and absorption spectroscopy are used to determine chemical composition. In emission spectroscopy the material whose composition is to be determined is used to produce a characteristic spectrum. Then, with a suitable optical system which includes a prism or diffraction grating, and a detector, the characteristic spectrum can be interpreted and the composition of the material determined.

The fluids or solids are placed in an arc, spark, gas discharge, or flame and heated to a point at which they emit their characteristic radiation if they do not radiate naturally. The recording of the spectra, the measurement of the intensity of the various lines, the various types of spectrographs, and the interpretation techniques are covered extensively in the literature.

ULTRAVIOLET TECHNIQUES

The concentration of an ultraviolet-absorbing material in a mixture may be determined fairly easily. The concentration is related directly to the amount of absorption from a beam of ultraviolet radiation that can be passed through the mixture. Thus, with an ultraviolet type of spectrophotometer, a number of absorbing components in a mixture can be identified simply on the basis of their patterns of absorption as a function of wave length. In principle, the ultraviolet transducer consists of a source of ultraviolet radiation, optical filters, a sample cell, a detector, and an output indicator.

The amount of transmission is determined by use of the ratio of the output signal ascertained with the cell filled to that obtained with the cell empty. The concentration can be determined from the known absorptivity of the substance by means of the Lambert-Beer law or by comparison with other samples whose concentration of known substances is well known. The ultraviolet sources that can be utilized in these transducers include tungsten lamps and arc discharge lamps that contain mercury, mercury-cadmuim, hydrogen, xenon, sodium, or other materials. Each type of lamp has its own characteristics, which must be selected for the particular problem.

REACTION PRODUCT DEVICES

Chemical composition can be determined by the measurement of a reaction product. First, the desired chemical reaction must be promoted, and then the reaction product must be measured to determine the presence and quantity of the constituents. A reagent-treated paper on a fabric tape can be suspended in the stream carrying the fluid. These devices are useful for monitoring exhausts and smoke, for determining the dusts, aerosols, and corrosive and toxic gases and vapors in polluted atmospheres, and for continuous monitoring, by use of moving tape, of the concentrations of specific components of gas mixtures. Concentrations can be determined from fractions of a part per million up to several percent.

A small area of impregnated tape is exposed to the gas sample, and a constituent in the sample then reacts with a reagent in the tape to form a reaction product. The reagent buried in the tape must be selected so that the reaction product will display a characteristic that can be detected such as a change in color, in electrical conductance, or in opacity. Tape can be moved continuously through the sample or individual samples can be taken. Suitable analytical instruments must be employed to determine the quantity of the characteristic change such as the change in conductance, etc., and relate it to the concentration.

With liquids, it is possible to employ a reaction that forms a dilute suspension. Or, the reaction product can cause a change in the color of the liquid. When solids are formed in the liquid because of the reaction, the amount of solid formed can be determined by photometry. From the change in the quantity of suspended particles in the liquid, the concentration can be measured. By application of the Tyndall effect, this technique can be employed for the quantitative determination of very small amounts of any material that is capable of reflecting light.

When color is changed, a differential optical transmission measurement can be utilized to determine the specific components of gas mixtures. The transmittance through a reference solution is compared with transmittance through a reagent solution in which the gas sample has passed under controlled conditions.

Concentrations can be determined also by a change in electrolytic conductance. A constant-temperature conductance cell is employed for the flowing electrolyte.

pH MEASUREMENTS

To determine the effective concentration of acids and bases in solution, the pH technique can be utilized. Special electrodes are employed to develop a voltage that is proportional to the hydrogen-ion concentration in the solution into which the electrodes are immersed. This technique of measuring pH refers only to the concentration of hydrogen ions that actually are dissociated in the solution and not to the total acidity or alkalinity.

Many techniques of pH measurement are available. Ordinarily they involve some type of an electrode, such as glass, antimony, or hydrogen. Also, a reference electrode is used, usually calomel or a silver — silver chloride unit. A potential measuring instrument, such as a vacuum tube voltmeter, indicates the output.

MASS SPECTROMETERS

When mass spectrometers are used to determine the chemical composition of a substance, the material to be analyzed is injected into some type of ionizing device. The resulting ions are separated according to their mass number by the combination of an electric acceleration field and a magnetic field. When an electrode is placed at a focal point, the resulting ion current is a function of the particular mass that is in focus at that station. By maintaining a constant electric acceleration force and by varying the magnetic field, one can sweep through a wide range of mass numbers.

If strength of the magnetic field, the charge, and the accelerating potential are known for a particular instrument the mass that is being received at the electrode may be determined. Various calibration techniques with known constituents may be used also. By recording the intensity of the ion current, one can determine the relative magnitude of the constituents. The most common angles used to bend the ion beam in the magnetic field are 60, 90, and 180 degrees. When ion source and the ion collector are suitably located, the ion beam is focused in a line at the collector electrode.

Special mass spectrometers can be designed and constructed for direct isotope-ratio measurements (or to determine ratios of other constituents), wherein two or possibly three electrodes are located at the various focus positions and the resulting ion currents are collected continuously at the various electrodes. The isotope ratio(s) can be measured directly by taking the ratio(s) of the output signals electrically.

X-RAY TECHNIQUES

X-rays may be used to determine the composition of certain materials and fluids on the basis of fluorescence, emission, absorption, and diffraction. It is possible to determine qualitatively and quantitatively the basic content of the constituents of complex mixtures in terms of the elements, and to determine exactly their atomic arrangement and spacing of the unit crystal. With diffraction techniques, one can determine the crystal sructure of metals and other materials at high temperatures and observe the phase changes as the crystal structure is modified by temperature variations.

ELECTRICAL CONDUCTIVITY TECHNIQUES

The ion concentrations in many solutions may be measured simply by electrical conductivity methods. The concentrations of various materials in simple water solutions can be determined with relative ease by conductivity techniques. The conductivity-concentration curve must be known in advance, or the calibration determined experimentally. Polarization effects can be reduced by using an alternating current rather than a direct current in the conductivity cell. Usually some type of AC Wheatstone bridge is used to determine the changes of conductivity in the conductivity cell. The cells are very simple in basic structure; they usually consist of two metal plates or electrodes that are fixed rigidly within an insulated chamber. Often platinum electrodes are used in pyrex glass cells.

OXYGEN ANALYZER

The concentration of oxygen in some cases may be measured by using the paramagnetic property of oxygen. The paramagnetic susceptibility of this component varies inversely as the square of the temperature of the gas and decreases rapidly as the temperature is increased. Usually the cell must have some type of temperature control so that the temperature can be held constant while the magnetic susceptibility is measured. The output signal is a function of the paramagnetic susceptibility of the gas volume and generally this signal is directly proportional to the oxygen concentration. For these measurements no paramagnetic substance other than oxygen can be present.

POLAROGRAPHY

Polarography is a method of chemical analysis based on comparative measurements of current-voltage curves that are obtained during electrolysis of a solution under specified conditions. Concentration polarization must occur at one electrode and a constant potential must exist at the other electrode. The various ions and molecules in solution can be identified and measured by this technique if they are susceptible to oxidation or reduction at the polarized indicator electrode by applied potentials in the neighborhood of a few volts. For many applications this technique is selective and accurate.

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Various isotopes can be identified separately according to their differing nuclear gyromagnetic constants, the basis for nuclear magnetic resonance spectroscopy. Just as a certain mass and electric charge are associated with each isotope, so also a spin or angular momentum is associated with each isotope.

With this method a magnet is employed whose field strength can be changed from essentially zero up to perhaps 10,000 gauss. A low-frequency radio transmitter supplies RF energy to a small transmitter coil, which is placed in the magnetic gap. A small receiver coil is located within the transmitter coil that surrounds the sample material to be tested. A sensitive radio receiver, which is tuned to the same frequency as that of the transmitter, is capable of amplifying any signal that might be induced in the receiver coil. Some type of indicator or recorder measures the presence of these signals, which can be related to the various individual isotopes.

BIBLIOGRAPHY

1. Baker, H. D., Ryder, E. A., and Baker, N. H., "Temperature Measurement in Engineering, Vol. 2," Wiley, New York, 1961.
2. Cerni, R. H., and Foster, L. E., "Instrumentation for Engineering Measurement," Wiley, New York, 1962.
3. Considine, D. M., "Process Instruments and Controls Handbook," McGraw-Hill, New York, 1957.
4. Cusick, C. F., "Flow Meter Engineering Handbook," Minneapolis-Honeywell Regulator Co., Philadelphia, 1961.
5. Eckman, D. P., "Industrial Instrumentation," Wiley, New York, 1950.
6. Gray, D. E., "American Institute of Physics Handbook," McGraw-Hill, New York, 1957.
7. Kallen, H. P., "Handbook of Instrumentation and Controls," McGraw-Hill, New York, 1961.
8. Ladenburg, R. W., Lewis, B., Pease, R. N., and Taylor, H. S., "Physical Measurements in Gas Dynamics and Combustion," Princeton University Press, Princeton, New Jersey, 1954.
9. LaJoy, M. H., "Industrial Automatic Controls," Prentice-Hall, Englewood Cliffs, New Jersey, 1954.
10. Lion, K. S., "Instrumentation in Scientific Research," McGraw-Hill, New York, 1959.
11. MacDonald, D. K. C., "Thermoelectricity," Wiley, New York, 1962.
12. Minnar, E. J., "ISA Transducer Compendium," Plenum Press, New York, 1963.
13. Monk, G. S., "Light," Dover, New York, 1963.
14. Stone, J. M., "Radiation and Optics," McGraw-Hill, New York, 1963.
15. Tyson, Jr., F. C., "Industrial Instrumentation," Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

SESSION 3: General

Chairman: Dr. Frank E. Gartrell
Assistant Director of Health
Division of Health and Safety
Tennessee Valley Authority

Dr. Peter K. Stein
Professor of Engineering
Arizona State University, Tempe

SUMMARY

Methods are presented for the classification of transducers and measurement systems. Transducers are classified by function, by input requirement, and by energy types. Measuring systems, the systems formed by combinations of transducers, are classified as unbalance systems, in which the output quantities are observed directly, and reference systems, in which output is compared to known quantity. The presentation incorporates general principles of measurement engineering, on which these classification systems are based....

CLASSIFICATION SYSTEMS FOR TRANSDUCERS AND MEASURING SYSTEMS

CLASSIFICATION OF TRANSDUCERS

INTRODUCTION

General

The process of measurement consists of transferring information from one component in the instrumentation chain to the next, until a final display on the readout instrument is obtained. This signal transfer from link to link will always correspond to a transfer of energy from one component to the next.

If energy is drawn from the source of the quantity to be measured, then the very phenomenon that is to be observed will be altered. When the criterion is that only a small amount of energy may be drawn from the source system in the process of measurement, then the word 'small' implies that:

The amount of energy drawn from the source system in the process of measurement must be small compared to the total amount of energy available in the source system.

Thus some knowledge of the source system must be available to the measurement engineer. In a temperature measurement in a small, cooling cup of water the available energy in the observed phenomenon is small and finite. In an atmospheric pressure measurement the reservoir of available energy is almost infinitely large.

In a good measurement system the necessary transfer of energy from the source system to the measuring system is minimum.

The measuring process always affects the phenomenon on which the measurement is made.

This is the first law of measurements.

Every measuring system, no matter what the quantity measured, consists of a chain of components that transform energy from one form into another. These transformations of energy may frequently occur in the same discipline. The process by which this energy transformation occurs is called transduction, and the components performing this operation are called transducers.

In every measurement chain one must distinguish between different transducer types depending on their function in the system. Furthermore, in order to identify, understand, and express the behavior of a transducer, it is necessary to define certain basic properties that serve to 'completely' specify the transducer. It finally becomes necessary to be able to combine transducers into a chain of measuring element links and to predict the behavior of the resulting measuring system on the basis of the known properties of the individual transducer elements.

Principles of Transducer Classification

Transducers can be (and have been) classified by a variety of different techniques. The predominant three are discussed in the following paragraphs.

1. By the function they perform in the measuring system, i.e., whether they are at the input or output of the measuring chain or whether they act as modifiers of the information to be transmitted. This manner of classifying transducers results in the following categories:
    a. Input or measuring transducers.
    b. Modifying transducers.
    c. Output or readout transducers.
2. By the input requirements of transducers. This division results in two basic classes of transducers:
    a. Self-generatinog (active) devices, which produce an energy output for a single energy input.
    b. Non-self-generating (passive or impedance-based) devices, which require two energy inputs in order to produce a single energy output.

    For each of these transducer classes it is possible to relate variables at each input and at the output in a convenient manner, rendering the system ready for mathematical operation.
3. By the energy types involved in the transduction process. If one recognizes eight forms of energy, it is possible to classify all conceivable transducers in a 'Transducer Space' containing 8-cubed, or 512, possible locations.

CLASSIFICATION BY FUNCTION

Measuring Transducers

The measuring transducer is the portion of the measuring system that transforms the quantity to be measured into another quantity more easily measured. Usually more than a single process of transduction is involved in this stage of a measuring system.

A thermocouple measuring temperature in a moving gas stream, for example:

a. The gas temperature is transformed into a related temperature of the thermocouple junction.

b. The temperature of the thermocouple junction is transformed into an electrical output in the form of voltage or current, depending on the instrumentation conditions.

A pressure-measuring device, for example:

a. The pressure is transformed into a force acting on a mechanical structure (diaphragm, bellows, etc.).

b. Some consequence of this force is then measured. For example, the displacement of the diaphragm is measured with a differential transformer; the strain in the diaphragm may be measured with a strain gage, etc.

An accelerometer, for example:

a. The acceleration, acting on a mass, is transformed into a force.

b. The force is transformed into an electrical charge by action on a piezo-electric material, or it may be transformed into an electrical resistance by means of a piezo-resistive material, etc.

Note that the three examples cited have one thing in common:

All the phenomena listed as items (a) are basic phenomena associated with the physical quantity to be measured. Thus the relationship between thermocouple temperature and temperature of the gas stream into which it is inserted is exclusively a heat transfer problem and really has nothing to do with measurement engineering as such, although this relationship must be known and understood by the measurement engineer who measures temperatures with thermocouples.

All the phenomena listed under items (b) are basic phenomena associated with the transduction process. There are almost countless phenomena in the physical world that respond in some way to temperature, force, etc. An entire branch of measurement engineering is devoted to the study of the physical laws that can be used as the bases for transducers. A glimpse into this field will ge given in a later section.

Modifying Transducers

Modifying transducers act on the output from the measuring transducers and may be divided into two varieties:

a. Intentional modification.

b. Parasitic modification.

Intentional Modifications. Intentional modification implies that the modification (or computing function) introduced by the transducer is at the desire of and under the control of the measurement engineer. Perhaps the most universal component in an instrumentation system which exemplifies the intentional modification approach is an amplifier (mechanical, hydraulic, pneumatic, electrical, etc.). An amplifier is the prime example of an intentional signal modifier: it produces at its output a signal that is a known and desired modification of its input. Examples of other desired modifications may be integration, differentiation, adding, filtering, etc.

Parasitic Modification. In the process of signal transmission, signal modifications may occur that are undesired and therefore parasitic in nature. Although the measurement engineer may be aware of these undesired modifications he may not be able to exercise full control over their presence and action. Such modifications are often called noise levels.

Prime examples of such modifying systems are transmission systems such as lead wires, switches, and slip rings. In the piezo-electric accelerometer example cited previously, the validity of relating the charge of a piezo-electric transducer to the input acceleration is entirely dependent on the choice of the cable that will connect the electrical charge with the portion of the measuring system that will measure the charge.

The most usual form of parasitic modification in transmission systems is caused by the resistance and capacitance of the lead wires connecting the measuring transducer with the intentional modifying transducer in the measuring chain.

In general, parasitic modifications may be multiplicative or additive in action, i.e., they may multiply or add to the desired signal.

Readout Transducers

The readout transducer transforms the modified signal into an indication that may be observed with human senses: a visual, audible, smellable, tasteable, touchable form. Examples of such readout transducers are galvanometers, dial indicators, direct-writing recorders, the color of a titration mixture, the smell of a chemical.

It is normally assumed that nothing follows a readout transducer. This assumption is somewhat erroneous and depends entirely on where the defined measuring system is cut off. If the system is cut off at the cathode ray oscilloscope tube, for example, then considerations of matching the optical properties of the light emanating from the tube to the optical characteristics of the human eye or photographic film do not enter into the picture. Carried to extremes, however, the system could be defined as going on through the human eye into the system within ourselves that transmits the external light stimulus to our brain and permits us to observe the phenomenon displayed on the cathode ray tube face.

CLASSIFICATION BY TRANSDUCER INPUT REQUIREMENTS

Introduction

The definition of a transducer as an energy conversion element immediately includes such commonplace transducers as:

Thermometers: The heat energy input results in mechanical displacement output against a controlled force in the glass tube.

Bourdon tubes: The pneumatic energy input results in mechanical rotation output against the restraining torque to the tube.

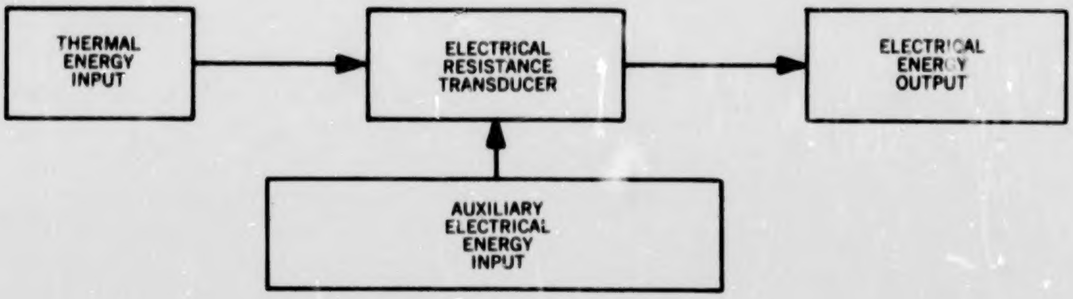
Other transducers do not seem to be covered by the energy concept as expounded so far.

Thermometer: Heat energy input results in change of electrical resistance.

But resistance (any impedance in fact) is not a form of energy nor is it a component of energy. It has been stated that any of the forms of impedance (resistance, capacitance, inductance) have no existence in themselves. To observe such elements one must either supply a current and observe a voltage or supply a voltage and observe a current.

Thus, to obtain an energy output from a resistance thermometer one must supply it not only with thermal energy input but also with electrical energy, so that the temperature-induced resistance change can be observed.

Thus a certain class of transducers requires two energy inputs to produce a single energy output. The additional energy input is often called auxiliary or biasing energy supply, or the minor or modulating input.


THERMAL ENERGY INPUT
ELECTRICAL RESISTANCE TRANSDUCER
ELECTRICAL ENERGY OUTPUT
AUXILIARY ELECTRICAL ENERGY INPUT

Transducers are classified as to whether the energy supplied by the unknown quantity to be measured (hereafter called the UQ) is sufficient to produce an energy output, or whether additional energy must be supplied to the transducer.

Self-Generating Transducers

Those transducer types for which the energy supplied by the phenomenon to be measured directly produces output energy are called self-generating transducers.

Examples:

1. Thermo-electricity: heat - electricity (thermocouples)
2. Mechanical levers: mechanical - mechanical energy
3. Piezo-electricity: mechanical force - electrical charge
4. Electrical generator: mechanical motion - electricity

Non-Self-Generating Transducers

Those transducers that require one or more auxiliary, minor, or biasing energy inputs to transform the action of the unknown phenomenon into an energy output are called non-self-generating, passive, or impedance-based transducers.

Examples:

1. Resistance-thermometer: thermal energy into electrical impedance
2. Resistance strain gage: mechanical energy into electrical impedance
3. Photoelasticity: mechanical energy into optical impedance
4. Inductance microphone: acoustic energy to magnetic impedance

Representation of Transducers

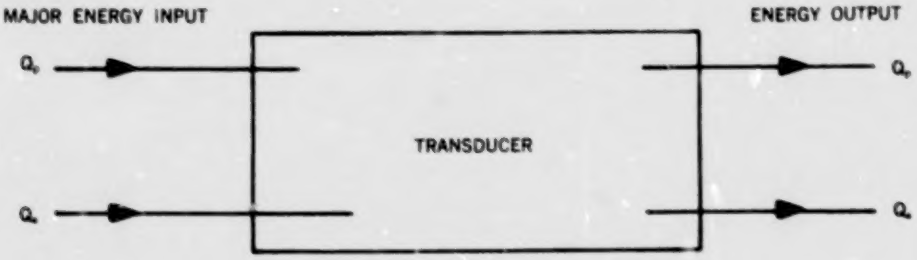
General: Since measurement implies the transfer of information through a transfer of energy, the definition of transducer has required the concept of an energy conversion device.

Energy normally consists of two co-existing physical quantities that are physically inseparable. Examples of such quantities, for which the product is energy, are:

Force and displacement	(mechanical energy)
Pressure and volume	(pneumatic-hydraulic energy)
Voltage and charge	(electrical energy)

One cannnot measure a force without permitting this force to go through some displacement. In so doing, the energy drawn from the system supplying this force and this displacement, could conceivably alter the force being measured. Neither is it possible to measure a displacement without force, although optical displacement measuring techniques could render such forces exceedingly small.

The representation of a transducer or of a measuring system consisting of a chain of transducers must, therefore, be in terms of energy flow — a concept requiring two inputs quantities at each input 'terminal' of the transducer. The representation and associated nomenclature for self-generating and non-self-generating transducers are shown below:



MINOR ENERGY INPUT

(exists only for non-self-generating transducers)

Q_p = primary quantity, i.e., the one to be observed.

Q_s = secondary quantity; the one that necessarily co-exists with Q_p.

Major input is that energy input containing the quantity to be observed.

Minor input, also called auxiliary, bias, and carrier input is that second energy input required for non-self-generating transducers to 'carry' the major-input-created impedance through the transducer to its output.

The only restriction on the choice of primary and secondary quantities is that dimensionally: $Q_p \times Q_s = \text{Energy}$.

It will be shown later that under certain special conditions this product could also be power.

The minor energy input: It has already been stated that the function of the minor energy input, especially in impedance-based transducers, is to 'carry' the major-input-created impedance-change to the output in the form of energy.

The properties, capabilities, and limitations of a measuring system are directly a function of the minor energy input used. These concepts could be further elaborated; only an indication of the possibilities involved is given below.

The classification of systems centers on two characteristics of the minor energy input, and one system design parameter. Any wave form can be used. The following are most frequently selected in commercial systems:

a. An invarient level (DC).

b. A sine wave.

c. A pulse train (square waves are considered as the special pulse train in which pulse duration and duration between pulses are equal).

The information being transmitted in the energy transfer 'carrier' process may be carried on any of the properties of the wave form used.

For a level input:

a. Amplitude

For a sine-wave input:

b. Amplitude (amplitude modulation, AM)
c. Frequency (frequency modulation, FM)
d. Phase (phase modulation, PM)

For a pulse train:

e. Amplitude (pulse amplitude modulation, PAM)
f. Frequency (pulse frequency modulation, PFM)
g. Position (pulse position modulation, PPM)
h. Duration (pulse duration modulation, PDM)
i. Width (pulse width modulation, PWM)
j. Presence or absence of pulses in a specified number of pulses (pulse code modulation, PCM)

The system performance will depend on whether the non-self-generating transducer requiring the minor input is located

a. as an input transducer
b. as a modifying transducer

Each of the 20 different systems possible in this classification alone will present different performance characteristics and can conceivably give 20 totally different answers in measurements of the same physical phenomenon.

Characterizing a Transducer (Self-Generating)

Mathematically, the behavior of a four-terminal system becomes defined when four specific coefficients for the system are known. This approach will be elaborated in Reference 5. The section that follows will approach the same problem from an intuitive way first.

To predict the behavior of a transducer, i.e., its output for any given input(s), one must establish at least three sets of relationships for self-generating-transducers.

1. Relations at the component input
2. Relations at the component output
3. Relations between component output and input

Relationships at the component input: There are always two quantities acting on the input of a transducer. The product of these input quantities will be either energy or power. Energy is the more fundamental form, but since energy is the time integral of power, it is usually accepted that measuring systems can be treated in terms of either the energy or the power transmitted through the system.

One of the two input quantities will be the one to be measured, the other quantity co-exists by physical necessity.

The primary quantity is the physical quantity to be measured. The secondary quantity is the physical quantity that co-exists with the primary quantity at the input.

The product of primary and secondary quantities will be the energy or power absorbed by the measuring system from the source system.

The ratio: $\text{acceptance ratio} = \dfrac{\text{primary quantity}}{\text{secondary quantity}} = A$

identifies the reaction of this system component with the one preceding it. This ratio permits the determination of how much the measuring system influences the physical process being measured.

The acceptance ratio is a complex number, mathematically, exhibiting both a magnitude and a phase angle (or a real and an imaginary component).

Each of the components of the acceptance ratio, i.e., its magnitude and its phase angle, depends on the frequency and the amplitude of the primary quantity.

Hence at the input alone, four characteristic equations or curves identify the reaction of the transducer to and on the preceding link:

$$|A| = f(a)$$
$$\angle A = f(a)$$
$$|A| = f(w)$$
$$\angle A = f(w)$$

where the symbol 'f' is the mathematical function symbol, 'a' connotes signal amplitude, and 'w' is radian frequency.

Relationships at the component output. There are always two quantities emerging from the output of a transducer. The product of these quantities will exhibit the dimensions of either power or energy. The dimensions of the product of the two quantities existing at the transducer input and at its output need not be the same. It is possible for the product of the input quantities to be energy in dimension and for the dimension of the product of the output quantities to be power. This condition prevails in all impedance-based transducers such as linear-motion potentiometers, strain gages, capacitive and inductive transducers, etc.

One of these quantities at the transducer output will be the one to be measured, i.e., transferred to the next link in the measurement chain. The other physical quantity co-exists of physical necessity. Once again:

The primary output quantity is the one to be measured. The secondary output quantity is the one that of necessity co-exists at the transducer output.

The product of primary and secondary quantities will give the power or energy that the transducer delivers at its output. The ratio

$$\text{emission ratio} = \frac{\text{primary output quantity}}{\text{secondary output quantity}} = E$$

identifies the reaction of this transducer with the one following it. This ratio permits the determination of how much the source system (transducer output) and the measuring system (input of the following transducer) interact and affect each other.

In general the emission ratio of a transducer will depend on the magnitude and on the frequency of the input quantities and will be a complex number, as was the case for the acceptance ratio. Hence, to characterize a link in the measurement chain, the measurement engineer must know:

$$|E| = f(a)$$
$$\angle E = f(a)$$

$$|E| = f(w)$$
$$\angle E = f(w)$$

Note: The concepts of acceptance and emission ratios are analogous to those of input and output impedance, but they are not equivalent. The impedance concepts can be derived from the acceptance and emission ratios, but not vice versa. The energy-based approach can be shown to be more fundamental and more universally applicable.

Relationships between output and input for the component. The relationships between transducer output and input may be defined as

$$\text{transfer ratio} = \frac{\text{primary output quantity}}{\text{primary input quantity}} = T$$

This ratio identifies the action of the transducer on the transfer process that the signal undergoes in passing through that component. The transfer ratio (also called gain, sensitivity, response ratio) will be a function of the magnitude and frequency of the transducer input. Thus again the ratio will be a complex number and not necessarily related in a linear manner to the system input.

Hence, to characterize the link in the measurement chain being considered, the measurement engineer must know:

$$|T| = f(a)$$
$$\angle T = f(a)$$
$$|T| = f(w)$$
$$\angle T = f(w)$$

General note on the ratios. When the statement is made that a ratio is a function of the amplitude or magnitude of the system input, the implication is that the relationship between input quantity and the ratio is nonlinear. In general it is possible to define ranges of input quantity for which the relationship is linear within certain limits of linearity (say one percent deviation from linearity). This range of inputs is then called the linear input range for the transducer.

The statement that a ratio is a function of signal frequency implies that the amplitude and the phase angle of the ratio may be both frequency dependent (i.e., the amplitude of the ratio or its absolute magnitude is frequency dependent and the phase angle between the numerator and the denominator or the direction of the vector representing the ratio may be frequency dependent).

Thus, to display the entire properties of each of these ratios the following must be known:

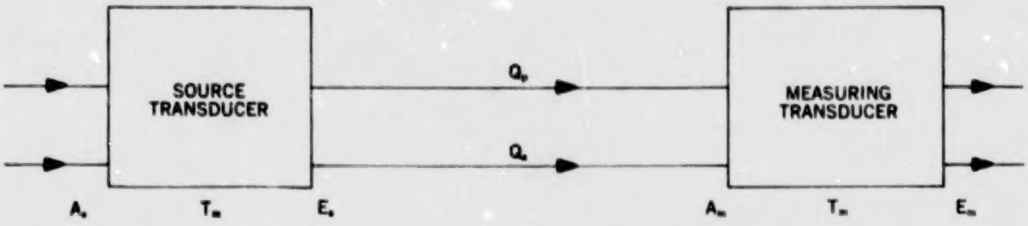
a. The magnitude of the ratio:
    its dependence on the magnitude of the input signal
    its dependence on the frequency of the input signal

b. The phase angle of the ratio:
    its dependence on the magnitude of the input signal
    its dependence on the frequency of the input signal

Interaction between transducers. In measuring systems, it is generally desired that a minimum of energy be transferred from the source system to the measuring system.

The source transducer is the transducer that immediately precedes the one being considered.

The two transducers are said to be isolated when they do not interact, i.e., when the transfer of energy from the source transducer to the measuring transducer is zero (or very, very small).



A measure of how isolated two transducers are is the ratio:

$$\text{Isolation ratio} = \frac{\text{acceptance ratio of measuring transducer } (A_m)}{A_m + \text{emission ratio of the source } (E_s)} = \frac{A_m}{A_m + E_s}$$

As this ratio approaches unity, the isolation between the transducers approaches perfection.

This ratio, too, is a complex number since it is a function of complex numbers, and it, too, will depend on amplitude and frequency of signal:

$$|I| = f(a)$$

$$\angle I = f(a)$$

$$|I| = f(w)$$

$$\angle I = f(w)$$

The isolation ratio also represents the portion of the primary quantity that is available at the source output under ideal conditions to the primary quantity available under the existing isolation conditions:

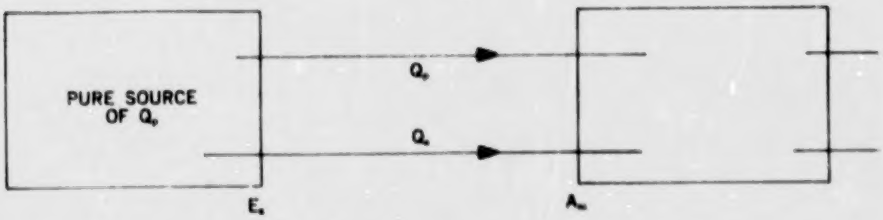
$$\text{isolation ratio} = \frac{\text{primary quantity obtained}}{\text{maximum available primary quantity}} = I$$

It can thus be said to be an efficiency indicator for the measuring system design.

CONCLUSION

If the basic characteristics of a component in the measurement chain are known, then the interaction and transfer characteristics of the component are known and it can be intelligently applied, selected, and used. If these characteristics are not available from the manufacturer of the transducer element, they must be determined experimentally or analytically; otherwise it is impossible to obtain valid data on purpose. One merely obtains data instead of making a valid measurement.

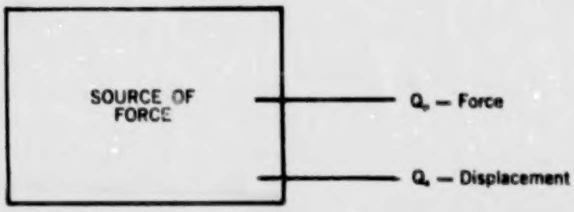
Characteristics of pure sources. A pure source of any physical quantity must have an emission ratio of zero. Only in that case can the isolation ratio between it and the elements which the source feeds be unity.


PURE SOURCE OF Q_p
Q_p
Q_s
E_s
A_m

$$\text{isolation ratio} = I = \frac{A_m}{E_s + A_m} = 1 \text{ only for } E_s = 0$$

Examples:

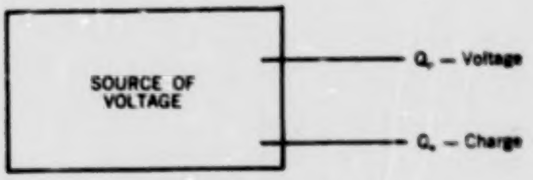
1. Pure Source of Force. A mass in the field of gravity acts as a pure source of force. It must be considered not as a mass that stores kinetic energy, but as the extreme example of a spring!


SOURCE OF FORCE
Q_p – Force
Q_s – Displacement

Given no restraints, the mass would be capable of undergoing an infinite displacement in order to apply its force (mg) to an object.
Hence its emission ratio: $Q_p/Q_s = mg/\infty = 0$

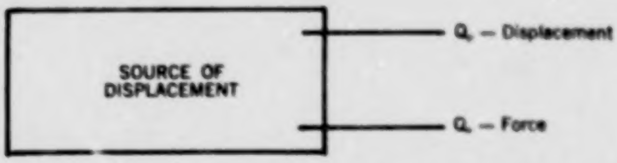
A mass in the field of gravity has an infinite spring constant and is to be considered as a pure source of force.

2. Pure Voltage Source


SOURCE OF VOLTAGE
Q_p – Voltage
Q_s – Charge

By definition, a constant voltage source must be capable of supplying any charge necessary to maintain a given voltage, and its emission ratio becomes $Q_p/Q_s = 0$

3. Pure Displacement Source


SOURCE OF DISPLACEMENT
Q_p – Displacement
Q_s – Force

A pure source of displacement must be capable of overcoming any force generated in the process.

CLASSIFICATION OF TRANSDUCERS BY ENERGY TYPES INVOLVED

The basic definition of a transducer implies that some input energy is converted into some output energy. That there may be more than a single type of input energy in order to arrive at an output in the form of energy was discussed for passive or non-self-generating transducers.

Furthermore, it was shown that energy is usually not the quantity that is to be measured; that usually the input and the output quantity to be measured are accompanied by secondary quantities such that the product of the two quantities at the input or at the output is the form of energy.

All properties of a transducer that describe its reaction with previous and following measuring system components, and that describe the action of the transducer on the input energy, were expressed in terms of the primary and secondary quantities at the input and output.

Basic Types of Energy Conversion

Energy can conveniently be divided into eight general classes, although the lines of distinction have grown less and less well defined over the years. For example, electromagnetic waves in certain frequencies are called light, in others they are called electromagnetic radiation; the point at which pressure fluctuations cease to contain acoustic energy and become mechanical is just as poorly defined. Examples of this type can be multiplied to encompass almost all the distinctive boundaries between the classical concepts of energy types.

For single-input energy — single-output energy transducers, i.e., active or self-generating transducers, the classification to be given may cover all the possibilities in types of transducing principles. Note that not all the conversions have yet been achieved, nor do they all form bases for transducers that have been achieved in the past.

For passive, non-self-generating transducers, requiring two or more energy inputs to produce a single energy output, the combination and permuation of these energies in threes yields a tremendously large variety of transducer types that can be envisioned.

A course in this field should undertake the study of non-self-generating transducers, since this field is the more general (and also more complex). The approach taken should be independent of the transducing principle used so that the principles presented could be applied to any of the possible transducing mechanisms.

Terminology

Some of the terminology applied to interactions between the energy classes is listed below:

Type of Energy	Adjectival and Combining Forms
Mechanics	Mechanical, mechano-, piezo-, -strictive, -elastic, -dynamic
Sound	Accoustic, -sonic (ultrasonic)
Heat	Thermal, thermo-
Light	Optical, photo-, spectro-, spectral, (infrared, ultraviolet), luminescent, phosphorescent

Electricity	Electro-, electric, electrical, electronic, galvano-, voltaic
Magnetism	Magneto-, magnetic, (paramagnetic, ferromagnetic, ferrimagnetic)
Chemistry	Chemical
Physics of the Nucleus	Nuclear, subatomic, nucleonics

A General Classification System: The Transducer Space Concept

For self-generating transducers with only one energy input and one energy output, the above concepts result in 64 possible transducer combinations when 8 forms of energy are considered.

In non-self-generating transducers the action of the major input, i.e., that of the physical quantity Q to be measured, creates a variation in passive property of the auxiliary energy system. Example: the temperature-induced electrical-resistance change in a resistance-thermometer. This passive aspect of an energy system has been called impedance and may be mechanical, electrical, thermal, etc.

To transform an impedance into an energy output it is necessary to apply a minor or biasing energy input as previously explained. Thus it becomes a simple matter to extend the two-dimensional "lattice" of self-generating transducer energy-conversion methods into a three-dimensional array with the auxiliary energy input as the third axis, as illustrated below, resulting in a transducer space.

This system will then permit the classification of any of the physical effects used in energy conversion, resulting in 512 transducer possibilities (in terms of energy-types combinations) when eight types of energy are distinguished, or 343 when only seven are recognized. (Dr. Lion in Ref. 2 combines acoustical with mechanical energy, for example.)

All transducers can now be classified by their location in the transducer space coordinate system:

Major Energy Input — Minor Energy Input — Energy Output

Examples:	Piezoelectric devices	503
	Thermocouples	803
	Electric resistance strain gages	533
	Electric resistance thermometer	833

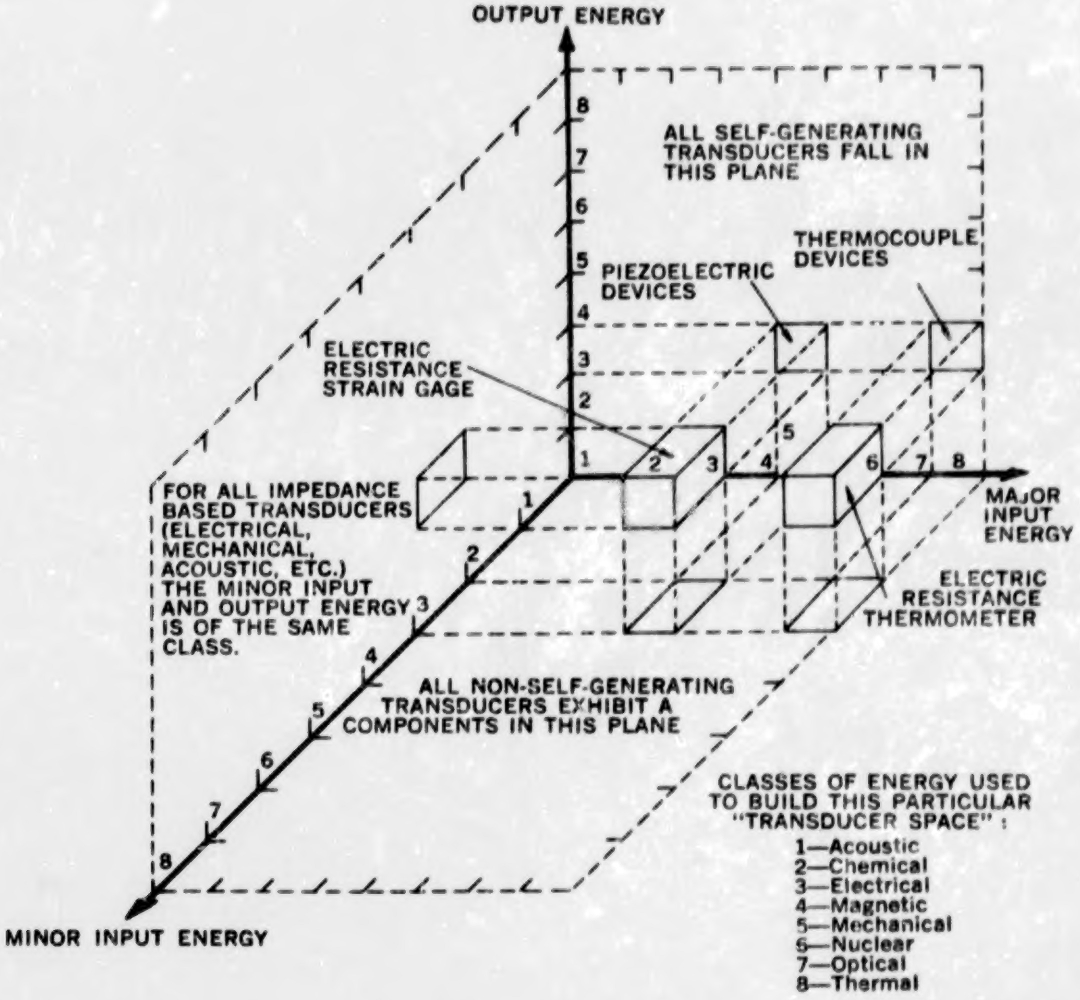
For all impedance-based transducers it is necessary that the minor energy input be of the same class as the energy output, so that the last two digits of the transducer classification will always be either "0x" or "xx."

An arbitrary number code is used for the arbitrarily selected classes of energy in the specific illustration of the general concept.

Utilization of the Transducing Possibilities

To study in detail the many possibilities that could be the bases for transducers would require a tremendous amount of work, and would resolve basically into a study of physics. What becomes important to the measurement engineer is to have the basic knowledge of these phenomena available, and a few good references at hand where

additional information can be found. Some of the most useful information in this field is contained in References 1-4.


OUTPUT ENERGY
8
7
6
5
4
3
2
1
ALL SELF-GENERATING TRANSDUCERS FALL IN THIS PLANE
THERMOCOUPLE DEVICES
PIEZOELECTRIC DEVICES
ELECTRIC RESISTANCE STRAIN GAGE
FOR ALL IMPEDANCE BASED TRANSDUCERS (ELECTRICAL, MECHANICAL, ACOUSTIC, ETC.) THE MINOR INPUT AND OUTPUT ENERGY IS OF THE SAME CLASS.
1 2 3 4 5 6 7 8
MAJOR INPUT ENERGY
ELECTRIC RESISTANCE THERMOMETER
1 2 3 4 5 6 7 8
ALL NON-SELF-GENERATING TRANSDUCERS EXHIBIT A COMPONENTS IN THIS PLANE
CLASSES OF ENERGY USED TO BUILD THIS PARTICULAR "TRANSDUCER SPACE" :
1—Acoustic
2—Chemical
3—Electrical
4—Magnetic
5—Mechanical
6—Nuclear
7—Optical
8—Thermal
MINOR INPUT ENERGY

The Transducer-Space Concept for Transducer Classification

CONCLUSION

The advantages of the foregoing definitions and representations become apparent in further study of measuring systems.

1. **They represent a measuring system for what it is, a system of information transfer through energy transfer.**
2. **They permit the application of a well-developed mathematical tool—the four- and six-terminal network theories.**
3. **They permit direct and simple classification of transducers by the energy types involved in the transduction process.**
4. **They permit the logical inclusion in the classification system of such information transfer methods as amplitude, frequency, pulse-width, pulse-duration, pulse-position, and pulse-code modulation techniques.**

CLASSIFICATION OF MEASURING SYSTEMS

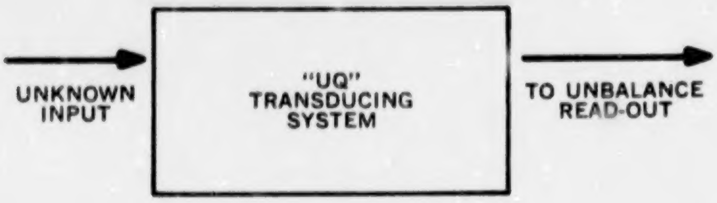
The methods by which a physical quantity may be measured are basically three, within two classifications:

1. Unbalance Systems
2. Reference Systems
    a. Based on unbalance techniques
    b. Based on null-balance techniques.

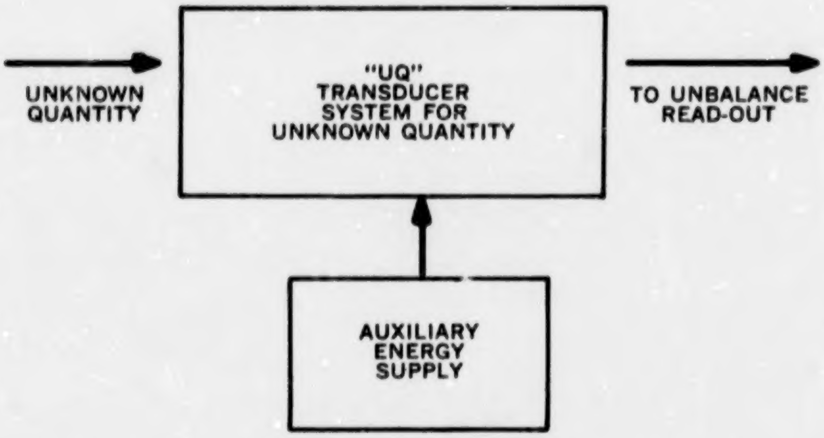
The capabilities and limitations of a measuring system depend to a very great extent on which of these three methods of measurement is used. Knowledge of the measurement method used in any commercial instrument applied to a test is vital. The most blatant measurement blunders are usually committed by users of transducers who either are not familiar with the fundamentals of their instrument or do not care to be. It is inevitable that the transducer itself, or the instrument, is blamed for the resulting erroneous data — never the user.

UNBALANCE SYSTEMS

In any measuring system, one set of input quantities (primary and secondary) produces one set of output quantities. In the unbalance measuring system the output quantities are observed directly and their magnitude is measured.


UNKNOWN INPUT
"UQ" TRANSDUCING SYSTEM
TO UNBALANCE READ-OUT

Unbalance System


UNKNOWN QUANTITY
"UQ" TRANSDUCER SYSTEM FOR UNKNOWN QUANTITY
TO UNBALANCE READ-OUT
AUXILIARY ENERGY SUPPLY

Unbalance System For Non-Self-Generating Transducers

Examples of unbalance systems are any meter on which needle deflection is taken as measure of the input quantity; the common spring scale; the speedometer on an automobile; the loudspeaker in a radio.

In unbalance systems the transfer characteristics of the measuring instrument are of vital importance in the interpretation of the measurement; so are the acceptance and emission characteristics. These ratios and their dependence on signal amplitude and frequency will govern system performance.

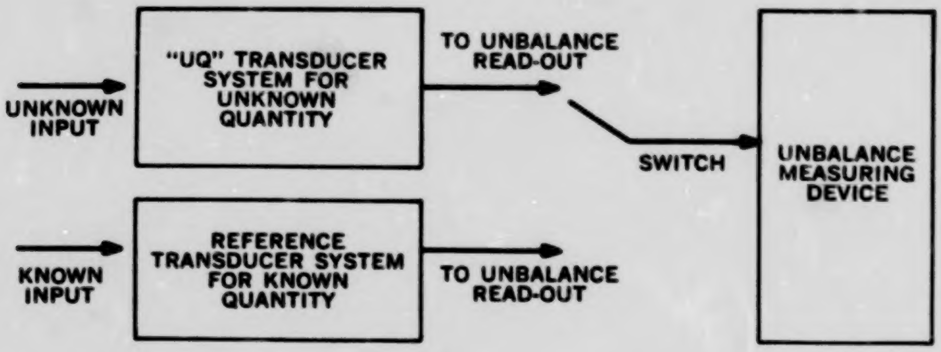
For a non-self-generating transducer the output will depend on both inputs for the unbalance system, and any variation in the auxiliary supply energy will influence system behavior.

REFERENCE SYSTEMS

In the reference systems of measurement, the transducer output is not observed, but it is compared to a known quantity. This known quantity is generated within the reference portion of the measuring system. The reference system output is varied until the unknown and known signals are observed to be equal or their difference is zero. Then the measurement is considered complete and the desired value is read from the reference system. The comparison may be of two types, and reference systems are distinguished as follows.

Based on the Unbalance Technique

The output of the UQ (unknown quantity) transducing system is compared with that of the reference transducing system by alternate switching between the outputs from the two systems to a common unbalance indicating device. The reference quantity will be called RQ.


UNKNOWN INPUT
"UQ" TRANSDUCER SYSTEM FOR UNKNOWN QUANTITY
TO UNBALANCE READ-OUT
SWITCH
UNBALANCE MEASURING DEVICE
KNOWN INPUT
REFERENCE TRANSDUCER SYSTEM FOR KNOWN QUANTITY
TO UNBALANCE READ-OUT

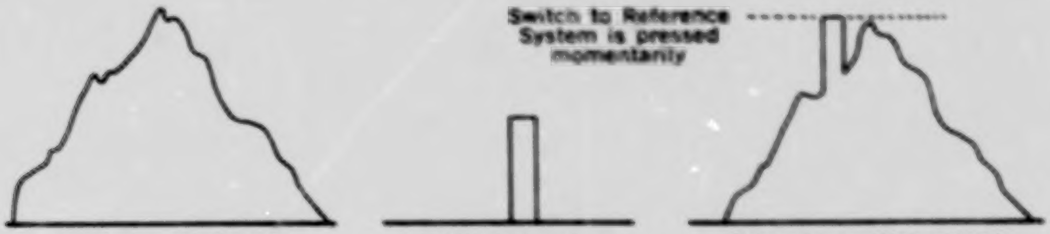
Reference Unbalance System

In such systems the measurement is independent of the linearity characteristics of the readout device since the unknown and known (reference) quantities are adjusted to be of the same amplitude. Thus any amplitude distortion in the system would be common to both signals and its effect on the reading eliminated.

The measurement would not, however, be independent of the frequency response of the readout instrument. When the reference system is switched into the readout device, the instrument "sees" essentially a pulse input. The frequency content of this reference signal may not be identical with that of the unknown signal so that frequency or phase distortions in the readout instrument would affect the unknown and known signals differently.

The measurement is also dependent on the isolation ratio between the measuring instrument and each of the sources (of the UQ and RQ). It can be shown that so long as the isolation ratio is high, even large differences in the emission ratios of the UQ and RQ sources will not materially affect system performance. The measurement also depends, of course, on all characteristics of the individual components in the UQ and the reference channels.

An example of the reference-unbalance measuring system is the Norwood Controls Pressure Indicator. The instrument is designed primarily for the measurement of dynamic steady-state pressures such as in internal combustion reciprocating engines. The unknown pressure wave may appear as in (a). By alternately switching to a reference circuit, which emits electrical signals equivalent to known pressures (b), one can adjust the output from the reference circuit to equal that of the unknown phenomenon. The required setting (c) on the reference circuit then gives the magnitude of the unknown signal.



Unknown steady-state dynamic pressure (a)

Reference system output switched momentarily on to output (b)

Reference system output adjusted to equal maximum unknown pressure (c)

Sample Unbalance Reference System

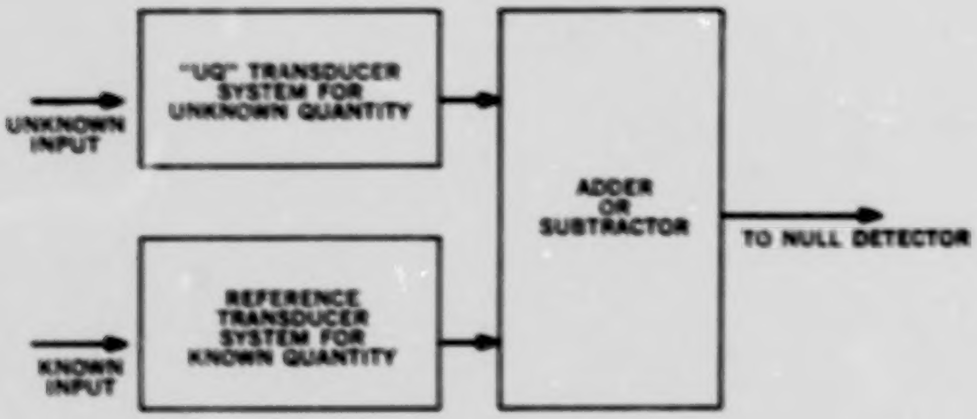
Special case of unbalance reference system — zero-reference. In a special type of unbalance reference system the basic reference may be zero output from the reference system, which would also correspond to zero output from the unknown system (since zero is zero).

In this case, it is merely required that periodically, the input to the readout detector be made zero; in an electrical system this is equivalent to a short-circuit, which is simple to achieve. In a typical instrument such as the Ellis Associates BA-13 (or BA-12) Bridge and Amplifier, the input to the amplifier is periodically short-circuited by means of an electromagnetically driven switch. In its closed position the switch short-circuits the amplifier input; in its open position the switch permits the signal to be measured to pass through the amplifier.

Based on Null-Balance Techniques

The UQ output is added to (or subtracted from) the reference system output. The reference system is then so adjusted that the *combined output* is zero. Then the reading on the reference system is equal to the unknown signal (in subtraction) or minus the unknown signal (in addition). Example: A mechanical balance for weighing.

Since under all conditions of data-taking the system output is maintained at zero, the system behavior is independent of the transfer characteristics or the acceptance ratio of the readout instrument. Thus input loading, linearity, frequency response, etc., of the readout instrument do not affect the accuracy of null-balance systems. On the other hand the rapidity with which the reference system can be adjusted to maintain zero output limits the frequency response of the system as a measuring system.


UNKNOWN INPUT
"UQ" TRANSDUCER SYSTEM FOR UNKNOWN QUANTITY
KNOWN INPUT
REFERENCE TRANSDUCER SYSTEM FOR KNOWN QUANTITY
ADDER OR SUBTRACTOR
TO NULL DETECTOR

Reference Null Balance System

Manual null balance will not accommodate signal frequencies over about ½ cps; mechanical servosystems may go to a few cps; electronic techniques can be used to extend the frequency response of such systems to higher limit.

Special case of non-self-generating transducers. For non-self-generating transducers, the unbalance reference system may be operated in one of two ways. The separate auxiliary energy supplies for the UQ and reference systems imply that any change in either auxiliary supply will affect the reading. It is then possible to take advantage of placing the transducer system outputs directly in series.

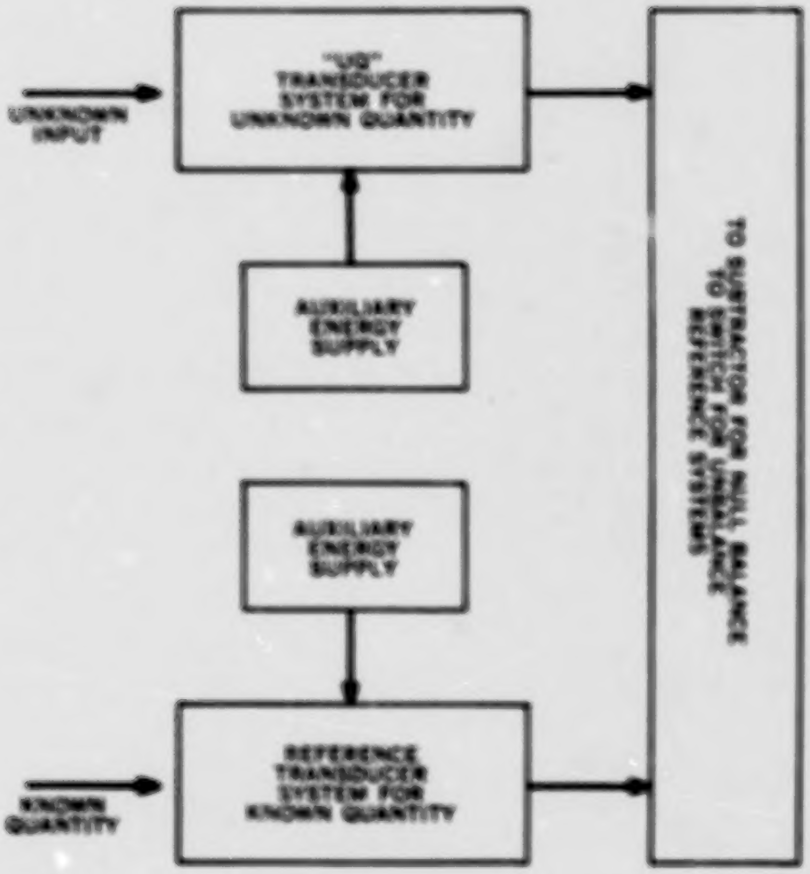
Where separate auxiliary energy supplies are used, the system is called a separate reference system. Where a common auxiliary supply is used, the system is called an integral reference system. The word "integral" means that the reference system is a part of the non-self-generating transducer and that both the transducer and the reference system are fed from a common auxiliary supply.

CONCLUSION

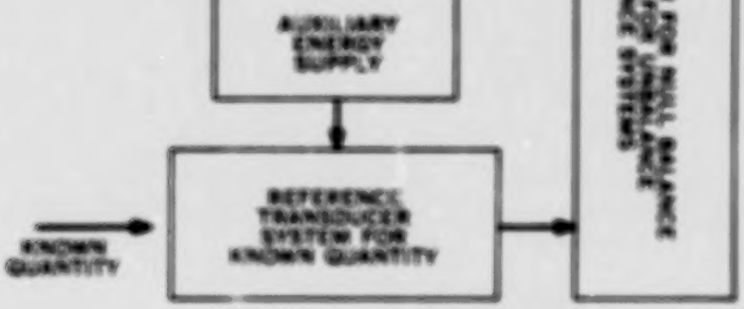
Depending on the measuring method selected, the characteristics of the readout transducer may or may not influence the measurement. Therefore, it is important to know both the measurement method and the readout instrument characteristics. Inherently the reference technique is capable of higher accuracy than the unbalance technique. Most precision measuring systems are based on a form of reference measurement.

A property of all integral reference systems is that a change in zero (or balance) results in a change in calibration; in other words, the very act of balancing the system affects its calibration or transfer ratio.

A property of all separate reference systems is that a change in calibration setting (transfer ratio) results in a change in zero; in other words, a zero-shift will result when the transfer ratio of the system is adjusted.


UNKNOWN INPUT
"UQ" TRANSDUCER SYSTEM FOR UNKNOWN QUANTITY
AUXILIARY ENERGY SUPPLY
AUXILIARY ENERGY SUPPLY
KNOWN QUANTITY
REFERENCE TRANSDUCER SYSTEM FOR KNOWN QUANTITY
TO SUBTRACTOR FOR NULL BALANCE
TO SWITCH FOR UNBALANCE
REFERENCE SYSTEMS

Reference System With Separate Auxiliary Energy Sources for Non-Self-Generating Transducers


UNKNOWN INPUT
"UQ" TRANSDUCER SYSTEM FOR UNKNOWN QUANTITY
AUXILIARY ENERGY SUPPLY
KNOWN QUANTITY
REFERENCE TRANSDUCER SYSTEM FOR KNOWN QUANTITY
TO SUBTRACTOR FOR NULL BALANCE
TO SWITCH FOR UNBALANCE
REFERENCE SYSTEMS

Reference System With Common Auxiliary Energy Sources for
Non-Self-Generating Transducers

These two properties are disadvantages of each system only under certain specific test conditions. There are tests in which only one or the other instrument should be used.

REFERENCES

1. Physical laws and their effects, C. F. Hix, R. P. Alley, John Wiley and Sons, 1958.
   Compiles a large number of physical laws, relating the different types of energy one to the other. A brief description of the law is given, an example of its application, an indication of the expected magnitudes, and one or two references. The effects are cross-indexed both by the proper name of the inventors or discoverers and by their scientific nomenclature.

2. Instrumentation in scientific research: input transducers, K. S. Lion, McGraw-Hill Book Company, 1959.
   Compiles a large number of physical laws which have actually been used as transducing principles for a large variety of measurements. Each section gives a brief description of the law, of actual transducers which have been made and used, based on this principle, and references in the literature applicable to the transducer are given. The book is organized for methodical presentation.

3. International critical tables, McGraw-Hill Book Company.
   Lists the physical laws and the various numerical coefficients which give the magnitudes of the various effects. This is a collection of much of man's experimental knowledge in all fields of science.

4. Searching the literature for transducer information: Part 1: A guide to the literature, J. Pearlstein, Report PB 161-320 from Office of Technical Services, Washington 25, D.C.

5. Measurement engineering, Peter K. Stein, Stein Engineering Services, Inc., 1962.
   A systematic survey and text on measurement engineering fundamentals.

Gerald C. Gill
Professor of Meteorology
University of Michigan, Ann Arbor

SUMMARY

To obtain valid data it is essential that attention be given to the following sequence of events: careful selection of the most suitable sensors and recorders for the parameter to be measured; proper installation; regular maintenance and servicing; and regular recalibration. An area often overlooked in a measuring system is the dynamic response of the sensors and of the recorder to fluctuating inputs. Grave errors in the recorded data may result from this oversight. Some fundamental relationships in this area are discussed and some useful curves reproduced.

DATA VALIDATION*

INTRODUCTION

We shall discuss the main factors that determine the accuracy and fidelity of recording a given variable. To specify the degree of accuracy and to maintain a high level of dependability, the investigator must consider the following factors:

1. A clear understanding of the principle of operation of the basic sensor and a knowledge of its dynamic response.
2. A general understanding of the principle of operation of the indicating or recording system.
3. Calibration of the system. Some instruments require only static calibration; others require dynamic calibration to determine the response of the system to a rapidly fluctuating variable.
4. Proper installation and use of the instruments.
5. Routine servicing.
6. Periodic maintenance.
7. Periodic calibration checks.
8. Alertness for small clues that may indicate errors developing in the system.

Before discussing these topics I will define some of the terms used in specifying the performance of instruments.

DEFINITION OF TERMS

The *sensitivity* of an instrument may be defined as the smallest change in the measured variable that causes a detectable change in the indication of the instrument. (Example: For a thermocouple recorder having a range of 100°C on 10-inch-wide chart paper, the sensitivity of a new and properly adjusted instrument might be

---

*Publication No. 79, Department of Meteorology and Oceanography, The University of Michigan. Research conducted under Research Grant #AP00233-01, from the Division of Air Pollution, Bureau of State Services, U. S. Public Health Service, and the sponsorship of the National Center for Atmospheric Research.

± 0.1°C, which corresponds to about 0.01 inch of pen movement. But the sensitivity might be as low as ± 1.0°C if the sliding contact were badly worn or the servo amplifier poorly adjusted.)

The *accuracy* of an instrument (including application of its calibration curve) is the precision with which the instrument will measure the variable in terms of internationally accepted units. (Example: the accuracy of the thermocouple recorder might be ± 0.5°C over the complete range when it is new and properly adjusted, but as poor as ± 2°C with worn sliding contacts and a weak servo amplifier.)

The term *speed of response* of an instrument is variously applied. Often it indicates the time required for the indicator or recorder to follow 90 percent of a sudden full-scale change in the measured variable; sometimes 99 percent of full scale. Sometimes the term indicates the time that elapses from the application of a sudden square-wave change until the recorder reading is steady. The term must always be defined.

For most sensors and recorders having a first-order response* the term *time constant* is much better, since it has only one meaning. Suppose the thermojunction of our thermocouple thermometer is suddenly transferred from an air stream at a constant temperature Θ_0 to a warmer air stream at constant temperature Θ_e (Figure 1). The thermo-junction will not instantly assume the new air temperature Θ_e but will change at a rate depending on the instantaneous temperature difference $(\Theta_e - \Theta)$, and will approach the new temperature asymptotically. *The time constant* is the period that is required for the temperature sensor (thermojunction) to respond to 63.2 percent $(1 - 1/e)$

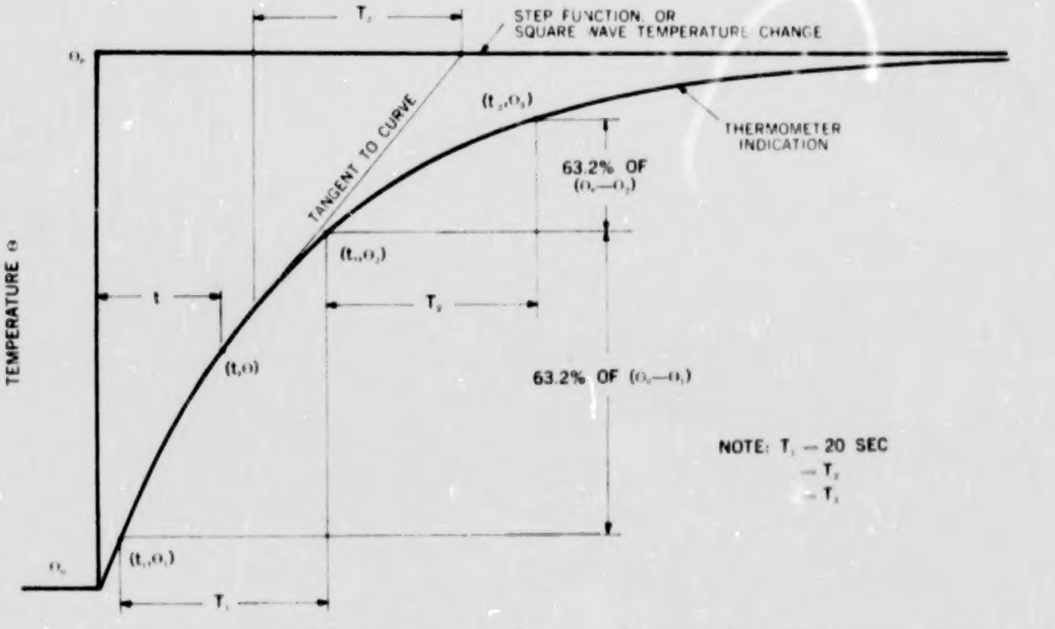


Figure 1 — Response of a Thermometer at Temperature Θ_0 and Time Constant T to a Sudden Change in the Environment (Step Function) to a New Temperature Θ_e

---

*See "Dynamic Response of Sensors."

of the stepwise change in temperature. (The significance of this constant is given in a succeeding section.)

For some sensors the term *distance constant* is more appropriate than the term *time constant*. For instance, when a three-cup anemometer is suddenly transferred from quiet air to a wind of 10 ft/sec the time constant might be 3.0 seconds, but if the same instrument were transferred from quiet air to a wind speed of 20 ft/sec the time constant would be only 1.5 seconds. The same amount of air (3.0 sec x 10 ft/sec = 30 ft; 1.5 sec x 20 ft/sec = 30 ft) will have passed in each case for the sensor to respond to 63.2 percent of the speed change. Thus the term *distance constant* is more appropriate for such a sensor. This is likewise true for propellor anemometers, propellor-type flow meters, etc. The distance constant of a sensor is the length of air column (or water column) required to cause it to respond to 63.2 percent of the square-wave change in speed.

In the calibration of an instrument the indications of the instrument are usually plotted against known values of the parameter for a number of points over the range of the instrument. Since these points generally do not yield exactly a straight line, instrument manufacturers usually draw a "best fit" straight line through the calibration points and specify the *linearity* as the maximum deviation of any points from the straight line. This linearity, often expressed as a percentage, refers to percentage of full scale deflection rather than percentage of the indication. (Example: For the thermocouple recorder, the linearity might be expressed as ± 0.5 percent. This would indicate a deviation of ± 0.5°C from true value over the complete range from 0° to 100°C). Some manufacturers specify that the straight line must pass through the zero of the recorder. In such cases the linearity then specifies the maximum deviation of any point from this straight line.

Most of you probably have conducted a static calibration of an instrument, and are familiar with the problems of making a reliable calibration; yet many of you are probably unfamiliar with the pitfalls of using such an instrument for measuring fairly fast fluctuations of the variable (ten fluctuations per second, or perhaps only one fluctuation per minute). Accordingly, it seems appropriate to outline the behavior of some general types of sensors with stepwise and with sinusoidal fluctuations of the variable being measured and to supply a set of curves and formulae that will be valuable in conducting dynamic calibrations of sensors and sensing systems.

DYNAMIC RESPONSE OF MEASURING SYSTEMS

DYNAMIC RESPONSE OF SENSORS

Sensor With First-Order Response (equation of forces being a first-order differential equation).

Consider a thermometer, initially at a temperature Θ_0, which is suddenly transferred into a moving air stream whose temperature is Θ_e (see Figure 1). Experiment shows that the indicating thermometer will approach the new temperature Θ_e asymptotically at a rate depending on the temperature difference $\Theta_e - \Theta$. This relationship may be expressed by the equation:

$$\frac{d\Theta}{dt} = \frac{\Theta_e - \Theta}{\lambda} \text{ (a first-order differential equation)} \qquad (1)$$

where Θ = instantaneous indication of thermal bulb at time t

Θ_e = temperature of new environment (assumed constant)

Θ_0 = initial temperature of thermometer

t = elapsed time (sec) after thermometer immersed in new environment

λ = constant, depending on shape and composition of thermometer bulb, and properties of new environment. (Note that λ has the dimensions of time in the equation.)

Solving this differential equation we get

$$(\Theta_e - \Theta) = (\Theta_e - \Theta_o)\, e^{-t/\lambda} \tag{2}$$

Now when time $t = \lambda$,

$$(\Theta_e - \Theta) = (\Theta_e - \Theta_o)\, e^{-\lambda/\lambda} = (\Theta_e - \Theta_o)\, e^{-1}$$

$$= \frac{(\Theta_e - \Theta_o)}{2.718} = 0.368\,(\Theta_e - \Theta_o) \tag{3}$$

that is, after the elapse of λ sec the instantaneous difference in temperature $(\Theta_e - \Theta)$ has been reduced to 36.8 percent of its original value, or, in time λ the thermometer will have responded to 63.2 percent of the initial temperature difference.

This constant λ having the dimensions of time is called the time constant T and corresponds to the elapsed time required after a sudden change in the environment temperature for the indicated temperature difference to be reduced to $\frac{1}{e}$ of its initial value.

Response of a first-order sensor to square-wave (step function) input. In time T seconds the sensor will have responded to 63.2 percent of the initial temperature difference. In the succeeding T seconds the sensor will have responded to 63.2 percent of the remaining temperature difference 0.368 $(\Theta_e - \Theta_o)$; that is, in 2 T seconds it will have responded to 86.5 percent of the initial temperature difference. Table 1 relates percentage response of the sensor to other values of elapsed time, measured in terms of the time constant.

Table 1 — Recovery of Sensor With First-Order Response After a Step-Function Input.

Recovery, %	50	63.2	90	95	99	99.5	99.9
Elapsed Time	0.7T	1.0T	2.3T	3.0T	4.6T	5.3T	6.9T

In determination of the time constant T by the method given above, other points on the curve beside Θ_o may be used. For instance, T_1 = elapsed time after t_1 for temperature indication to reach Θ_2, where Θ_2 is determined by equation. $(\Theta_e - \Theta_2) - 0.368\,(\Theta_e - \Theta_1)$. T may also be determined by drawing a tangent line to the curve and noting elapsed time T_3 where it cuts line Θ_e (Figure 1).

The unique value of the term 'time constant' is shown in Figure 2, which illustrates what happens when the same thermojunction sensor has been immersed in an air stream at the same speed as before and at constant temperature Θ_o, when at time $t = o$ the temperature is raised at a constant rate. The temperature sensor does not immediately respond to this constant rate of temperature rise but takes about 40 seconds to reach this rate. The time constant T is the lag time T_4 or T_5.

For a given air speed (or water speed) and a given temperature sensor the time constants T_1, T_2, T_3, T_4, and T_5 should be the same within a few percent.

(Note - The time constant for a thermometer exposed in an air stream at a certain speed is about 60 times greater than it is for the same thermometer exposed in a water stream at the same speed.)

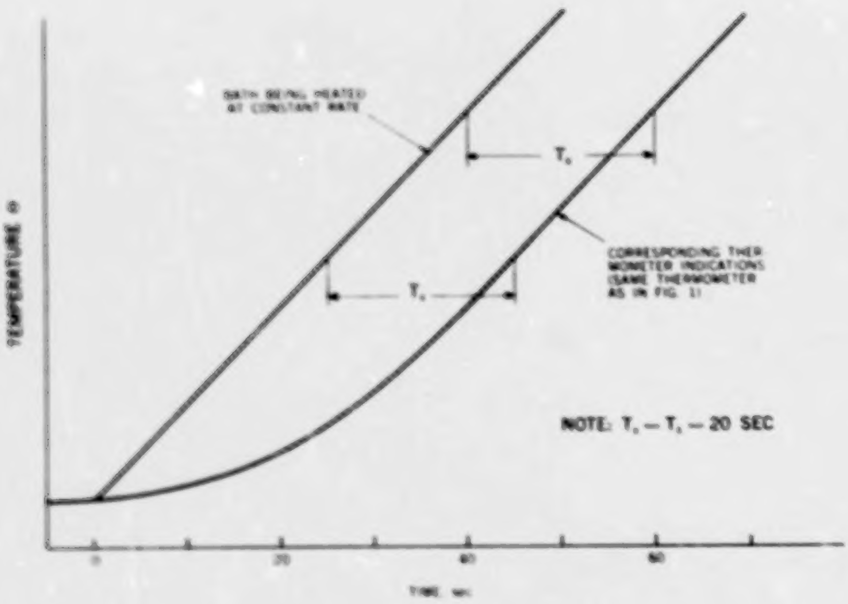


Figure 2 — Response of a Thermometer (With Time Constant T) That is Exposed in an Air Stream That is Suddenly Heated at a Constant Rate

Response of a first-order sensor to a sinusoidally fluctating input. The accuracy of indications of the amplitude of a sinusoidally fluctuating input is given by the following equations:

(1) Amplitude ratio for a single-capacity system (e.g., bare resistance wire, or butt-welded thermojunction):

$$\frac{x}{x_o} = \frac{1}{\sqrt{1 + (\omega T)^2}} \tag{4}$$

$$\text{or} \quad \frac{T}{P} = \frac{\sqrt{(x_o/x)^2 - 1}}{2\pi} \tag{5}$$

(2) Amplitude ratio for a double-capacity system (bulb in well):

$$\frac{x}{x_o} = \frac{1}{\sqrt{1 + (\omega T_1)^2}\ \sqrt{1 + (\omega T_2)^2}} \tag{6}$$

where x = indicated amplitude

x_o = actual amplitude

ω = angular velocity (radians/sec) $= 2\pi f = \frac{2\pi}{P}$

f = frequency of fluctuation

P = period of fluctuation $= \frac{1}{f}$

T = time constant

T_1 and T_2 are time constants of bulb alone and well alone.

Figure 3 is a graphical representation of Equation 5 relating the time constant T and the period P of the cycle to the amplitude ratio $\frac{x}{x_o}$ of the sensor. (Example:

Suppose the time constant of a thermometer in a wind of 10 mph were 100 seconds and that sinusoidal air temperature fluctuations of ± 5° F were occurring at 5-minute periods ($P = 300$ sec). The ratio $\frac{T}{P} = \frac{100}{300} = 0.33$. From the graph the amplitude ratio would be 0.43. The sensor would show only 43 percent of the true temperature fluctuation.) Thus by knowing the time constant of the sensor and the period of the fluctuations, we can quickly specify the fidelity response of the temperature sensor. If we want the temperature sensor to respond to 90 percent of the temperature fluctuations, the ratio $\frac{T}{P}$ must be .075 or less; that is, the time constant of the sensor could not be more than 7.5 percent of the shortest period of fluctuations the sensor is to follow.

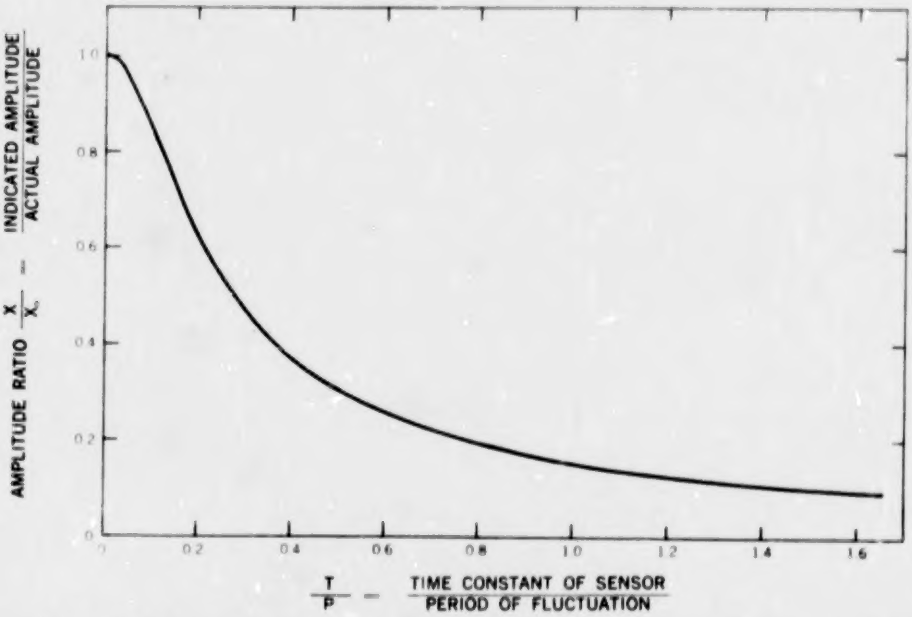


Figure 3 — Relationship Between the Time Constant T of a Temperature Sensor, the Period P of a Sinusoidal Temperature Fluctuation of the Environment, and the Fidelity of Recording this Fluctuation.

$$\left[\frac{T}{P} = \frac{\sqrt{(X/X_0)^2 - 1}}{2\pi}\right]$$

Figure 4 illustrates the response of several average temperature sensors to a sinusoidally fluctuating air temperature having a period of 300 seconds.* The effects of time constant on amplitude ratio and phase shift are clearly demonstrated.

The following formula (experimentally determined) relates the diameter of cylindrical metal temperature sensors, the air flow rate, and the time constant:

$$T = 6000\, d^{1.34}\, v^{-0.40} \tag{7}$$

where T = time constant (sec)
d = diameter of cylinder (inches)
v = air speed (ft/ min).

Note that the time constant is roughly proportional to the square root of the wind speed.

---

* Courtesy of E. W. Jensen and K. C. Kiesling, Eastman Kodak Co., "Response of Thermal Bulbs," I. S. A. Instrument Maintenance Clinic, Buffalo, New York, Sept. 15-17, 1950.

GPO 814-105-4

Sensors with first-order response. Essentially all temperature-sensing instruments have first-order response, e.g., mercury-in-glass thermometers, gas thermometers, resistance thermometers, etc.

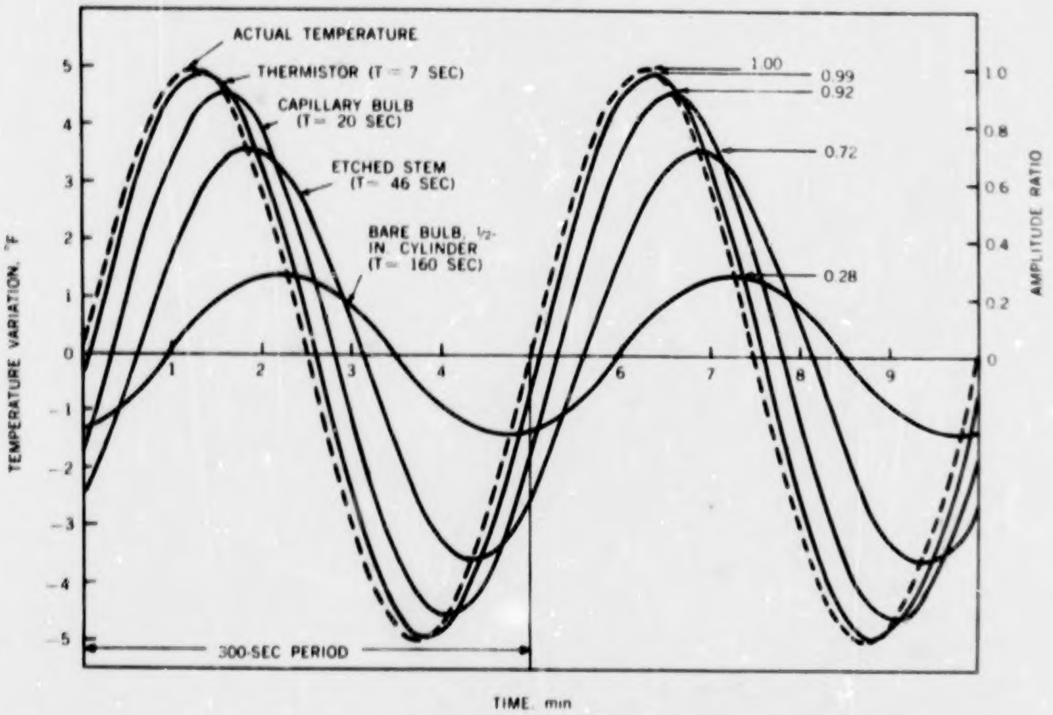


Figure 4 — Response of Typical Temperature Sensors to a Cycling Air Temperature of 5°F in Amplitude and of 300 Sec Period

Many flow-measuring sensors have first-order response, e.g., cup, propellor, and hot wire anemometers; cup and propellor water-speed-measuring sensors and turbine types of sensors.

It should be noted that a first-order sensor never indicates a larger change in the measured variable than the true change, even with a step-function input.

Sensors With Second-Order Response (equation of forces being a second-order differential equation).

Response of a second-order sensor to a square-wave (step function) input. In the electric circuit of Figure 5 after switch S has been closed for some time the galvanometer reading G will have become steady at some value, say A degrees. If at time t = O switch S is opened, the galvanometer coil will quickly start turning toward its zero position; overshoot this value by maybe 60 percent; reverse direction; and again overshoot, executing a simple harmonic oscillation of decreasing amplitude, as shown in Figure 6. (Here the sensor does indicate a larger change than the true change — larger by a factor of 60 percent — thus differing markedly from sensors with first-order response.) The equation of forces is:

$$I \frac{d^2\Theta}{dt^2} + C \frac{d\Theta}{dt} + K\Theta = O \qquad (8)$$

where I = moment of inertia of the coil suspension system

C = damping constant (primarily self-induced electromagnetic damping; secondarily air damping)

K = spring constant

Θ = angular deflection (measured from rest position)

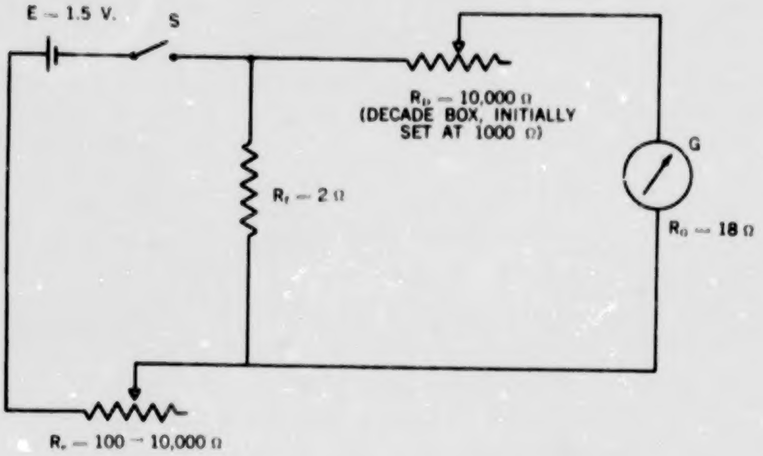

E = 1.5 V.
S
R_D = 10,000 Ω
(DECADE BOX, INITIALLY SET AT 1000 Ω)
G
R_f = 2 Ω
R_G = 18 Ω
R_s = 100 → 10,000 Ω

Figure 5 — An Electric Circuit to Determine the Critical Damping Resistance of a Galvanometer or Indicating Meter

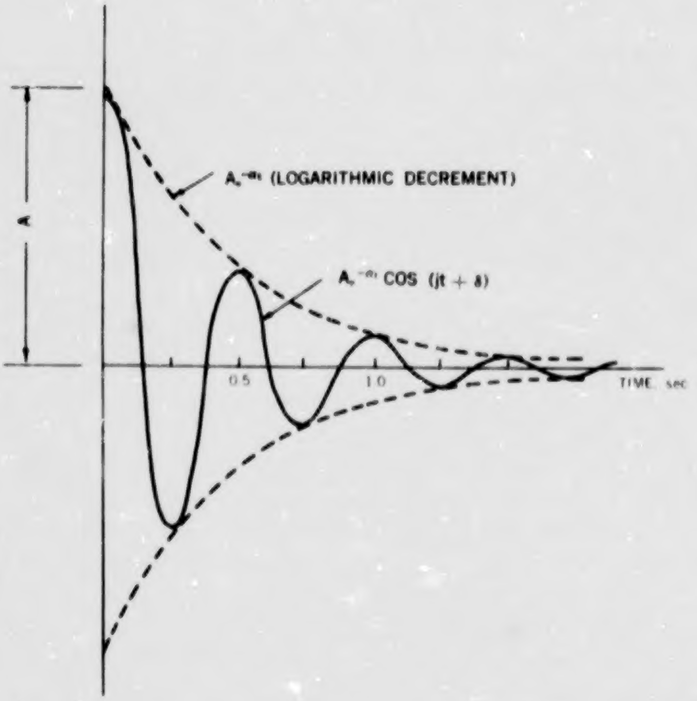


Figure 6 — Typical Galvanometer Decay Curve (h = 0.16)

A general solution to this equation (when the decay curve is similar to that of Figure 6) is:

$$\Theta = Ae^{-\alpha t} \cos (\gamma t + \delta)$$

where A = initial displacement from zero (9)

$$\alpha = \frac{C}{2I}$$

$$\beta = \sqrt{\frac{K}{I}}$$

$$\gamma = \sqrt{\beta^2 - \alpha^2}, \qquad \cos\delta = \sqrt{1 - h^2}$$

As mentioned previously, this decay curve is a simple harmonic motion ($\cos \gamma t$) of decreasing amplitude with envelope defined by the dashed curve $Ae^{-\alpha t}$.

If the electrical resistance R_d were decreased, the damping coefficient would be increased and the decay curve would show fewer oscillations, each of decreased amplitude. For a certain value of ($R_d + R_f$) the instrument deflection returns to zero in a minimum of time without any overshoot. This condition is known as *critical damping* and is shown in Figure 7 by curve 6, labelled h = 1.0. Other values of the damping ratio h from 0.0 to 3.0 are shown in Figure 7.

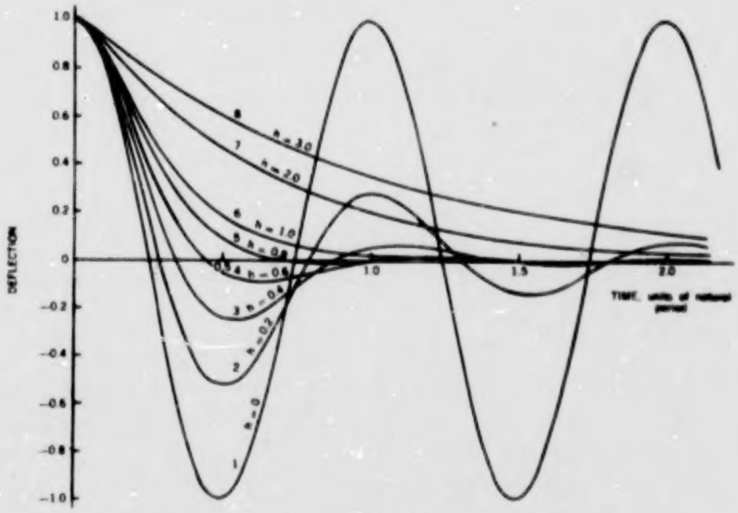


Figure 7 — Damped Oscillations — Galvanometer Decay Curves for Damping Ratios of h = 0.0 to h = 3.0

In the decay curve of Figure 6, the galvanometer has a damping ratio h = 0.16, and a damped frequency of 2 cps (damped period t_d = 0.5 sec). In the circuit of Figure 5 with R_d = 1000 ohms, if a sinusoidal voltage of constant amplitude but of varying frequency were applied across resistance R, the galvanometer would swing back and forth at the same frequency as the input signal; when the input frequency approached 2.0 cps, the galvanometer would resonate and show amplitude fluctuations up to 3.0 times the true amplitude! Thus the dynamic response of the sensor can greatly distort the true form of the input signal.

Figure 7 shows the decay curves of a galvanometer whose damping ratio h has been varied in steps from h = 0.0 to h = 3.0. Where h = 0.0 the galvanometer would execute simple harmonic motion without decreasing amplitude indefinitely. (This would represent a frictionless galvanometer without electrical or air damping, a theoretical case.) Time is given in units of the natural period (t_n) of the galvanometer. Note that as the damping ratio increases the "damped period" t_d increases. The relationship between the "damped period" and the damping ratio is given by

$$t_n = \sqrt{1 - h^2} \times t_d \tag{10}$$

or, $$f_n = f_d \div \sqrt{1 - h^2} \tag{10'}$$

where t_n = natural period of oscillation of galvanometer
f_n = natural frequency of oscillation of galvanometer
t_d = damped period
h = damping ratio

Note that for h = 0.2 the first overshoot ≐ 52 percent of the initial displacement; the second overshoot ≐ 52 percent of first overshoot, etc.

The damping ratio can be determined from the decay curve of the sensor by the use of the following equation:

$$h = \frac{\log_{10} \dfrac{A_o}{A_n}}{\sqrt{1.862n^2 + \left(\log_{10} \dfrac{A_o}{A_n}\right)^2}} \tag{11}$$

in which A_o is the first considered amplitude or displacement measured from the rest position and A_n is the amplitude on the nth succeeding swing past the rest position.

As an alternative to solving this equation for each test, Figure 8 relates the damping ratio h with the first overshoot after release, (that is, in Equation (11) A_o = initial displacement; A_1 = first overshoot; n = 1). (Example: In Figure 7, curve 2, the first overshoot is approximately 52½ percent. Referring to Figure 8, with an abscissa of 52½ percent the damping factor h = 0.20, which agrees with the value of h given for curve 2.)

Response of a second-order sensor to a sinusoidally fluctuating input. In most applications we are not concerned primarily with the response of the sensor to a step-function (square-wave) input, but rather with its response to a sine-wave input. Figure 9 shows the dynamic response of sensors with damping ratios from h = 0.1 to h = 1.0. This graph shows that for a galvanometer with a damping ratio of 0.2, if the impressed frequency f_1 were 0.5 that of the natural frequency f_n of the galvanometer (that is, $f_1/f_n = 0.5$), the galvanometer would indicate sinusoidal fluctuations 1.25 times that of true; when $f_1/f_n = 0.95$, the galvanometer would show fluctuations up to 2.50 times that of true (amplitude ratio = 2.50); and at $f_1/f_n = 2.0$, the amplitude ratio would be only 0.32, or ⅓ that of true. Thus for a galvanometer having a damping ratio of 0.2 for good fidelity in indicating, the ratio $f_1/f_n \ngtr 0.2$, or, the impressed frequency should never exceed 20 percent of the natural frequency of the galvanometer. This is a serious limitation on the use of the system because often one cannot limit the input frequency. If the galvanometer were damped to h = 0.64, the galvanometer would record the true input amplitude (within ± 2%) for all input frequencies where the ratio $f_1/f_p \leq 0.6$; and would record input signals less than true for all higher input frequencies. This value *h = 0.64 is the desirable damping ratio for most* applications, and manufacturers generally specify the input circuit resistance needed to achieve this ratio. (Note that if the galvanometer were critically damped, h = 1.0, and high-fidelity recording were desired (within ± 2% of true), the input frequency should never exceed 15 percent of the natural frequency of the galvanometer, a very serious limitation on the system.)

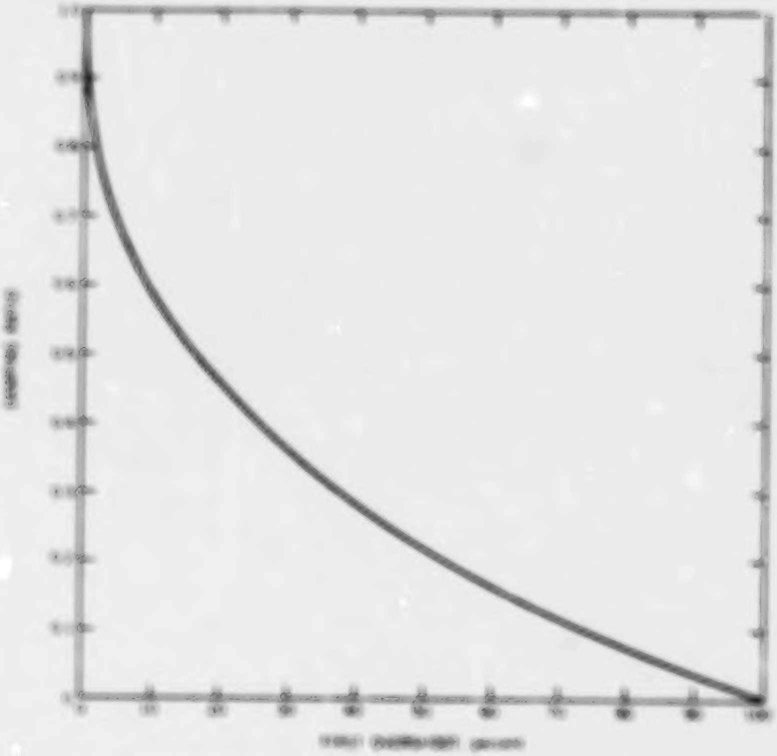


Figure 8 — Relationship Between the First Overshoot and the Damping Ratio of a Galvanometer

Most galvanometers have a damping ratio of 0.2 or less in a high-resistance circuit. This would be true of ammeters, voltmeters, etc., were they not provided with air or electro-magnetic damping. Such meters on open circuit usually have values of h 0.6 to 0.7, depending on their intended use.

Thus to specify the accuracy of recording fluctuating input frequencies, one must know the damping ratio of the sensor, its natural frequency, and the range of frequencies of the input signal.

Methods of increasing the damping of sensors with second-order response. As already mentioned, for best dynamic response (least distortion), the sensor should have a damping ratio of approximately 0.64 or higher. For electric meters this damping can be arranged by decreasing the input circuit resistance, or providing air or oil damping, or both.

In a wind vane with a particular area of vane, little can be done to change the damping factor C, but the moment of inertia I of the vane can sometimes be reduced without reducing the torque constant K in Equation (8). For most commercial wind vanes h = 0.1 to 0.3. Thus Figure 9 shows that such wind vanes will resonate with gusts of certain wave length, showing fluctuations up to 2 to 3 times the true angular fluctuations.

By reducing the moment of inertia of the vanes (using very light plastics) damping ratios as high as 0.6 have now been obtained. In this way sensors have been made that do not erroneously magnify the angular movements of the wind.

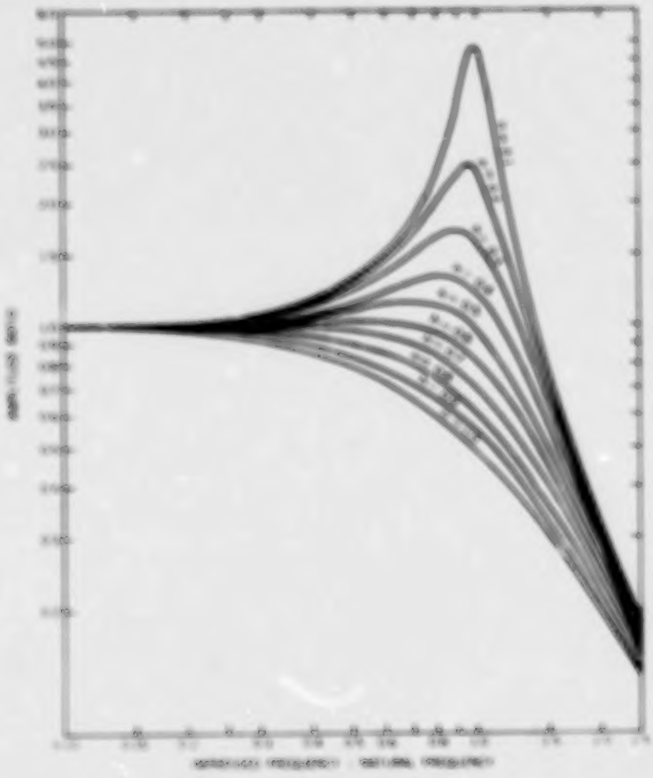


Figure 9 — Relationship Between the Damping Ratio of a Galvanometer (or Voltmeter) and Its Dynamic Response to Sinusoidal Input Voltages of Constant Amplitude but Varying Frequency

Sensors with second-order response. Electric meters generally are in this category. Fortunately, manufacturers usually provide their meters with damping ratios of 0.6 or higher or indicate the circuit resistance that should be used to attain this value.

Flow meters that incorporate a tapered tube with a float have second-order response.

Force sensors in which the force is balanced against a spring (either longitudinal extension or angular rotation) usually have second-order response.

In all such cases the dynamic response of the sensor must be known or measured if the accuracy of recording is to be specified.

Note — Sensors with first-order response are essentially special cases of second-order response, in which h = 1.0, that is, the sensor is critically damped.

DYNAMIC RESPONSE OF INDICATING METERS AND RECORDERS

Aronson* lists ten basic types of recorders. Of these, the galvanometer types and the null-balance types probably account for over 75 percent of the analogue recorders in routine use.

---

* M. H. Aronson, "Basic Types of Recorders," Recorder Manual, 1962 Edition, Instruments Publishing Company, Inc.

Galvanometer Types (second-order response)

For most indicators (voltmeters, ammeters, etc.) some mechanical damping is incorporated in the meter circuit to bring the damping ratio in the region of 0.4 to 0.7. If the damping ratio is not given in the specifications of such instruments, it can easily be obtained by the use of a circuit similar to Figure 5.

Most direct-writing galvanometer recorders (such as Esterline-Angus and Texas Instrument) incorporate some internal damping to permit movement of the recorder without shorting of the terminals. The manufacturer usually specifies the value of this damping factor or provides a set of typical response curves of the instrument for step function input of varying internal resistance. Fast-response recorders (such as Sanborn, Brush, etc.) are not damped in this way, but again the manufacturer supplies the dynamic response of the sensor. With all of these recorders the manufacturer's recommended circuit resistance should be used for best dynamic response.

Null-Balance Potentiometer Recorders

Quoting from Aronson

> Null-balance recorders are servo-operated devices that are generally referred to as potentiometers. The basic advantages of the null-balance potentiometer are (1) high sensitivity, down to microvolt signals, and (2) independence of lead length. The sensitivity is realized by the inherent amplification in the servo system; independence of lead length is realized by cancelling out the input signal so that no signal flows at balance.
>
> These two basic advantages are gained at the expense of response speed; potentiometers cannot operate at speeds faster than about ¼ second full-scale pen travel, limiting the response to signals of less than 1 cps. However, at these frequencies the potentiometer principle opens up vast areas for recording.

Most null-balance recorders have first-order response, and therefore, shows no overshoot for a stepwise input. In the specifications of such instruments it is usual to state the time for the recorder to indicate 90 percent or 99 percent deflection after application of the step function. If either of these is given, it is a simple matter to obtain the time constant of the recorder by reference to Table 1.

Null-balance potentiometers are made in many different forms. Some types could have a second-order response; these provide enough damping that negligible or no overshoot occurs with the step-function input. Accordingly, first-order response can be expected from most null-balance recorders.

DYNAMIC RESPONSE OF SENSORS PLUS RECORDERS

First-Order-Response Sensor; First-Order-Response Recorder

With this combination, where the time constants of both the sensor and the recorder are known, the dynamic response of the system may be obtained simply by use of Equation (6). Such a system will never over indicate fluctuations.

First-Order-Response Sensor; Second-Order-Response Recorder

If the input sensor is slow in response relative to the recorder, then the recorder damping ratio will approach 0.64 independently of circuit resistances. But if the response of the sensor is very fast relative to the recorder, the sensor will follow sinusoidal fluctuations in the variable without overshoot; to avoid overindication of the input variable the circuit resistance of the system must be designed so that the recorder damping ratio is 0.60 or greater. (Example 1: Consider a thermocouple sensor having a time constant of 20 seconds and a galvanometer recorder having a natural frequency of 1.0 cps (e.g., Texas 0-1 ma recorder). From Figure 3, for the sensor to respond to 90 percent of the sinusoidal fluctuations $T/P = 0.08$, $P = T/0.08 = 250$ seconds, or the impressed frequency to the recorder $= 1/250$ cps. Referring to Figure 9, $\frac{\text{impressed frequency}}{\text{natural frequency}} = \frac{1/250}{1} = \frac{1}{250}$, which is offscale to the left. Thus the recorder would faithfully follow the fluctuations without distortion whether $h = 0.1$ or $h = 1.0$ or higher. Example 2: Consider a hot wire anemometer, whose time constant is 0.1 second, connected to the same recorder. From Figure 3, for the sensor to respond to 95 percent of speed fluctuations, $T/P \doteq 0.05$, $P = 0.1/0.05$ sec $= 2$ sec, or, the impressed frequency $= 0.5$ cps. In Figure 9 when $\frac{\text{impressed frequency}}{\text{natural frequency}} = \frac{0.5}{1.0} = 0.5$ the damping ratio should be 0.60 or greater if no overindication of wind speeds is to be recorded. If the circuit resistance were such that $h = 0.60$ for this recorder, the sensor and recorder would be almost ideally matched for good dynamic operation.)

Other Combinations

The techniques just discussed will also apply to these combinations: second-order-response sensor with first-order-response recorder; and second-order-response sensor with second-order-response recorder.

STATIC CALIBRATION OF MEASURING SYSTEM

The term calibration is used to relate the indications of an instrument to internationally accepted units of measurement. Some recorders, which are built for a specific purpose, are equipped with special charts that indicate the desired units directly. Others operate with universal charts, whose values must be converted. In both cases one subjects the sensor to a series of known values of the variable, notes the corresponding deflections, and plots the calibration curve. For recorders with special chart rolls that read units directly, the calibration curve relates errors in indication to values of the parameter over the instrument range. For recorders with universal charts, the calibration sheet usually indicates the true value of the variable versus divisions on the chart roll, and the set of points is joined by a smoooth curve.

If the instrument system has been in prolonged operation, the system should be calibrated before it is adjusted or serviced. The calibration then applies to the readings that were taken during previous operation. For future use of the recording system, the basic sensor and the recorder should be carefully checked before the second calibration is made. For null-balance potentiometer recorders, one should check the freedom of operation of the writing system; the absence of backlash in the writing pen; the absence of end play in the chart-drive roller; performance of the servo-drive system (as shown by the pen returning to within ± 1/100 inch when deflected to right or left); condition of the battery, if any; and operation, adjustment, and lubrication of all other moving parts of the system. Galvanometer recorders require fewer adjustments but should be oiled and checked for proper adjustment before a calibration run.

Full calibration is usualy done in a laboratory, but sometimes it is desirable to calibrate the sensors and recorder in the field. As an illustration, if the recorder is normally mounted on a wall that is 30° F warmer or cooler than normal room temperature, the calibration should be conducted with the instrument in place. If large diurnal temperature fluctuations occur at the recorder site, the calibration should include tests to determine any errors due to these fluctuations. If accuracy of calibration within ± 1% of range is required, one may use four or five calibration points to cover the full scale of the instrument. If the accuracy of ± 0.3% of full range is desired, the system should be checked for at least ten values. For such accuracies the error involved in measuring the value of the parameter must be significantly less than the precision desired in the calibration. (Example: If a temperature system is to be calibrated within ± 0.3° C, the actual temperature must be measured with an accuracy of at least ± 0.1° C.) Both the sensor and the recorder must be allowed to come to an equilibrium position.

Multi-point recorders must be checked for any internal errors caused by the switching circuit. Usually a full calibration is not required for each multi-pen position, but it is well to record a particular sensor on each of the multi-points in succession to determine whether differential heating of the terminal block or of the switching circuits causes any error. (Example: In a thermocouple temperature recorder of a supposedly reputable manufacturer, the terminal block was located near one end of the amplifier system, causing differential heating of the block. This caused a progressive error in the circuits, so that temperature indications at the circuit nearest the terminal block differed by as much as 2° C from those at the junctions at the opposite end. When the terminal block was moved to a point remote from any heat source, this error was reduced to 0.1° C.)

If nonidentical sensors are used on a multi-point recorder, each sensor should be calibrated with the recorder.

Present-day good quality recording systems usually require full calibrations only once a year, if check calibrations are made periodically. One might check a multi-point temperature recorder by immersing a temperature sensor in a well-stirred bath of carbon tetrachloride at bimonthly periods. The temperature of the bath would be measured by a mercury or alcohol thermometer whose calibration was known. If the instrument were still within the previous limits of error of the system, full calibration would not be required.

Calibration checks can be built into some systems. (Example: in a resistance thermometer system one could use one or more precision resistances that would be automatically sampled at regular intervals, either with a strip-chart recorder or in a punch card system.)

MISCELLANEOUS FACTORS AFFECTING THE ACCURACY OF A RECORDING SYSTEM

Proper installation of the sensors and the recorder is imperative for accurate, reliable measurements. Many excellent instruments in which the basic sensor was poorly located have yielded observations that were almost valueless. (Example 1: If a wind-direction-measuring instrument were placed at the recommended height of 30 feet above ground but located within 60 feet and in the lea of a building 40 feet tall, the recorded wind directions would not represent the general area but would only indicate the eddies around the building. Example 2: If an accurate thermocouple system were installed with the thermojunctions at selected heights above ground but exposed to direct solar radiation without radiation shields and without artificial aspiration, the temperature readings could be several degrees high on calm sunny days and several degrees low on clear calm nights, even though the calibration was accurate within ± 0.3° C over the complete range.)

For reliable observations over periods of months the recorder should be checked daily at a specified time to insure proper operation and to place time marks on the chart roll. Daily maintenance should include a check for proper inking, for proper indication of the time, and for general system operation. Whenever chart rolls are changed, the operator should place enough data on the starting end of the roll to distinguish it positively from any other charts that might be used in the system complex. For instance, the wind direction chart at one level on a tower might be identified as follows: "Wind direction, 256 ft level, Charlevoix, on 0803 EST, Feb. 4/62, John Doe." A similar entry placed on the end of the roll thus completely identifies the chart records.

Generally such instrument systems should be thoroughly checked at about quarterly intervals. This check should include routine checks on the basic sensors, oiling and servicing where appropriate, and full maintenance and servicing of the recording system. Some inking systems require only very occasional cleaning of the pen points and the ink wells, say, at quarterly intervals. Other systems will require thorough monthly flushing of the ink wells and weekly cleaning of the pens for consistent fine-line traces. A careful maintenance and servicing routine can yield good records 99 percent of the time, whereas moderately careless servicing may yield less than 50 percent.

An alert technician detects trouble before it becomes serious, takes corrective action, and thus avoids loss of continuous records. He should report any variation from normal operation to his superior. Servicing personnel should be encouraged to obtain continuous, reliable records nearly 100 percent of the time; 95 percent is poor; less than 90 percent may make the record almost unusable. Most researchers are frustrated when even 1 hour of data is missing in a month; 35 hours (5 percent) of missing data constitutes a very serious loss.

The collection of accurate, reliable data is no accident. It is possible only through proper selection and installation of the measuring system, adequate maintenance and servicing, careful calibration at regular intervals (interspersed with routine checks), and continuous alertness for possible errors or failures in the recording system.

Richard S. Green
Chief, Basic Data Branch
Division of Water Supply and Pollution Control
U. S. Public Health Service, Washington, D.C.

SUMMARY

As pertinent data become more widely needed by groups and agencies involved in air and water pollution control, greater importance must be placed on uniformity of sampling and analytical procedures and on the accessibility of reliable data from all sources. Wherever acquired data are likely to have lasting value, serious thought should be given to some system of storage and retrieval through which potential users of the data can obtain the information they need in a usuable form at minimum cost and with reasonable speed and can be assured that all reliable information is included and that all extraneous information is excluded. To be workable, the system must be comprehensive, flexible, and simple. The Division of Water Supply and Pollution Control of the Public Health Service has devised such a system for storing and retrieving data for water quality control.

THE STORAGE AND RETRIEVAL OF DATA FOR WATER QUALITY CONTROL — A SUMMARY

INTRODUCTION

Collecting data and putting data to use costs a great deal of money, as those in this room know probably better than most others. A relatively simple chemical analysis of a water sample, with no unusual determinations, for example, costs even the most efficient laboratory $30 to $50 to run. We ought to get the most out of every dollar spent for such work, and we can help to reach this goal if all reliable data are made easily available to those needing them.

Special studies of all sorts produce large amounts of data, but little thought can usually be given to possible use of the data by others, for different purposes. Because of variations in objectives and requirements — resulting in different quality parameters, levels of concentration, period of sampling and the like — the data in original reports cannot easily be presented in a uniform format. Moreover, a large body of valuable data never appears in print at all, but remains in inaccessible files until discarded.

Whenever data are likely to have lasting value, we should give serious thought to some system of storage and retrieval wherein potential users of the data:

1. Can obtain the information they need *in the form they need it.*
2. Will be assured that all reliable data, wherever produced, are included in the material requested, and that all areas of interest are covered.
3. Will not be bothered with data they do *not* need.
4. Will get this service at minimum cost and with reasonable speed.

Any such system, to be workable, must possess three important characteristics. It must be:

1. *Comprehensive* — have the ability to handle all possible quality and related parameters, both those now in use and those that may be significant in the future.
2. *Flexible* — be able to take into account geographic and environmental differences.
3. *Simple* — be relatively easy to use and within reasonable cost range.

Many problems are involved in the design of a system of this scope. Most have been encountered and solved in a procedure for storing and retrieving data for water quality control devised for general use in the operations of the Division of Water Supply and Pollution Control of the Public Health Service. A description of the elements of that system, which has been named STORET, will bring out many of the basic principles involved.

ORIGINS

This system was developed from ideas brought together in a brief informal conference held in the Public Health Service about 2 years ago. The thoughts and suggestions of a few state officials who had been concerned about this problem were contributed by PHS personnel familiar with their views. Operating procedures in Indiana, New York State, and Pennsylvania were especially helpful. I should like to pay special tribute to the skill, tenacity, and patience of Assistant Sanitary Engineer Clarence Tutwiler, of our staff, who has been responsible for the electronic computer programming required in this system. He has adjusted and readjusted the storage and retrieval procedures several times as the full potential of the original concepts brought out in the 1961 meeting have become apparent.

SCOPE OF SYSTEM

The size and complexity of this data handling problem dictates the use of electronic computers with their great storage capacity and ready access to selected items. Two major concepts are being applied uniformly throughout the country, regardless of the computer equipment or programming technique utilized. These entail:

1. A single procedure for the identification of point locations pertinent to the data, whether they are water quality sampling points, points of waste discharge or of water intake, or any other locations for which data are to be secured.
2. A uniform coding system for the identification of specific parameters of water quality or other items of interest, such as data on flow, precipitation, and the like.

In the time available I shall only summarize these two concepts. The text of the full paper describes the storage and retrieval procedures in detail.[1]

Location Code. The location code permits the retrieval of data in the hydrologic order that is desirable for studies of basin problems. Since the system is complete and "open ended," it is possible to identify any point on any stream by it. Once a data point or station is given its proper location code, this "label" remains with that data point indefinitely. The location code is not, at present, completely adapted to points in estuarial waters where interlocking channels cannot easily be fitted into the concept, nor to large open water bodies. It was deemed unwise, however, to delay the application of other features of the plan on this account. Some form of coordinate location system will probably be used for such points.[2]

Parameter Code. One of the most troublesome features in the handling of data, particularly water quality data, for wide geographic areas is that we find, first, a large number of different kinds of data of concern to us. Also, in any given type of measurement there are usually widely varying limits to the values reported from place to place: Chlorides in New England streams are usually low, whereas those in some streams in the Southwest may be very high. If we provide a fixed field size on a punch card to accommodate the maximum values to be reported, we waste columns when less than that number is used. Furthermore, if we make room on a fixed field card for several parameters of data, not all of which are reported in each use of that card type, we waste still more columns. The availability of magnetic tape data storage enables us to overcome these difficulties. The parameter code adopted will handle up to 100,000 different parameters. Blocks of numbers have been assigned for specific parameter groups, leaving wide areas unfilled for additional future determinations and related data of interest.[3]

In this system, the code for any individual parameter is the same wherever the data are secured, and any potential user of the stored information can call for specific kinds of data through use of the proper code numbers. A special comment is necessary with respect to the handling of biological data. The number of individual entities here, i.e., species of organisms or fish, that may need to be reported and retrieved from storage for a given station, the need to record supplementary data about each entity, and the need to store and retrieve these data in associated groups require a slightly different method of storage and retrieval than that proposed for all other types of data. The procedures for this modification of STORET to handle biological data are now being developed.

The problem of variable number of digits in the value for any given parameter is handled by limiting the reported value to four significant figures, with the decimal point coded as the applicable exponent of 10.

Statistical Analyses of Data. Early in the design of this system it was decided that no attempt would be made to build in any procedures for statistical analysis of the stored data since this would only complicate the job of storage and retrieval. Since the output from the system can be in the form of data on tapes in prearranged order as requested by the user, however, subsequent statistical processing is a simple matter.

Status of the System. The programming of all routines involved in this data storage and retrieval system has been completed. In preparation for full-scale use of the system, index coding and stream mileage measurements are being undertaken at the present time by several of the comprehensive river basin projects within the Division of Water Supply and Pollution Control.

REFERENCES

1. The Storage and Retrieval of Data for Water Quality Control. Richard S. Green. PHS Publ. No. .... (in press).
2. Location Coding for the STORET System, Basic Data Branch, Division of Water Supply and Pollution Control, Public Health Service.
3. Parameter Code List for the STORET System, Basic Data Branch, Division of Water Supply and Pollution Control, Public Health Service.

DISCUSSION

Mr. Ransell expressed concern about duplication of effort in gathering water quality data and asked whether the storage and retrieval system described would be a repository for all valid data. Mr. Green indicated that the Public Health Service needs a national system for its own operations. The system described is capable of accepting data from all sources.

Mr. Ransell inquired about the organization required for handling input to the central system from various groups. Mr. Green explained that within the Public Health Service each comprehensive project is responsible for handling its own data. Therefore, most of these projects will probably operate within the system as self-contained units. It is expected, however, that the data from one part of the country will be useful and assimilable in other locations. Discussions have been held with State representatives and others about the broader use of a system (not necessarily this system) for handling data on water quality and related variable data throughout the country. Several participants of these discussions feel that a truly national system of handling water data should be formulated. The Public Health Service is interested in this and would be willing to contribute towards this goal.

Mr. Ransell cited current duplication of data-collecting effort as a reason for establishing one national collection system. Mr. Green pointed out that the storage and retrieval system described here is designed to accept data from various agencies and that ready availability of this information should be a stimulus toward reducing duplication.

Dr. Gartrell asked whether this storage and retrieval system, which will incorporate data from the National Water Quality Network, could be used for handling the vast amount of water quality data collected by the TVA. Mr. Green explained that although the built-in mechanical features of the system would permit this, the budgetary and administrative problems would have to be worked out. The Conference of State Sanitary Engineers and others are interested in this whole problem area, and future discussions are anticipated. Criteria for determining whether specific data will be stored or discarded must be formulated by all participating agencies.

SESSION 4: Measurements of Air Environment

Chairman: **George J. Taylor**
Supervisory Air Sanitation Engineer
California State Department of Public Health

John S. Nader
Chief, Physical Research and Development Section
Division of Air Pollution
U. S. Public Health Service, Cincinnati

SUMMARY

Two major automated data acquisition systems are now being used in the United States for air quality measurements. These systems, operated by the Los Angeles County Air Pollution Control District and by the U.S. Public Health Service (Continuous Air Monitoring Program), are reviewed in detail; plans for automated data handling by the California State Health Department are discussed briefly. Design and operation of these systems are reviewed in terms of sampling, detection, recording, data validation, and data display.

DATA ACQUISITION SYSTEMS IN AIR QUALITY

Automatic data collection and data processing in air quality had an early start in the Air Monitoring Network of the Los Angeles County Air Pollution Control District (LACAPCD), which was initiated in 1948. The U.S. Public Health Service began its Continuous Air Monitoring Program (CAMP) in September 1961. Currently, the California State Health Department is implementing its Berkeley station with automatic digital recording equipment as a pilot study toward a uniformly automated network of stations throughout the state.

This review is primarily directed to the first two of these air quality data acquisition systems, with respect to their major operations and their component elements. These networks are essentially the only air quality data acquisition systems that are fully automated and encompass the various operations in an environmental measurement system from the sampling of the ambient air to the display of validated data on pollutant concentrations in an accepted tabulation.

An air quality data acquisition system can be shown (Figure 1) to consist of the following basic operations:

1. Sampling
2. Detection
3. Data Recording
4. Data Validation
5. Data Display

The first two operations usually are performed by an integral and automatic instrument for sampling and analysis. Often the data recording is partially included as an analog recorder that produces a strip chart recording, which normally must be converted either manually or instrumentally to digital data to be compatible with subsequent data handling operations. Consequently, it is convenient to consider the first two operations as the components of one major subsystem, which generates the analog data for various parameters under study. The last three operations may be viewed as another major subsystem, which acts on the analog data to produce an acceptable display of information for subsequent operations of data analysis, interpretation, and drawing of conclusions.

Although the Los Angeles County Network takes historical precedence, the air quality data acquisition system of the USPHS is discussed first for convenience in presentation. The USPHS system is the more completely automated, particularly with respect to the recording of the analog data in digital form; in addition, all stations are equipped with identical equipment.

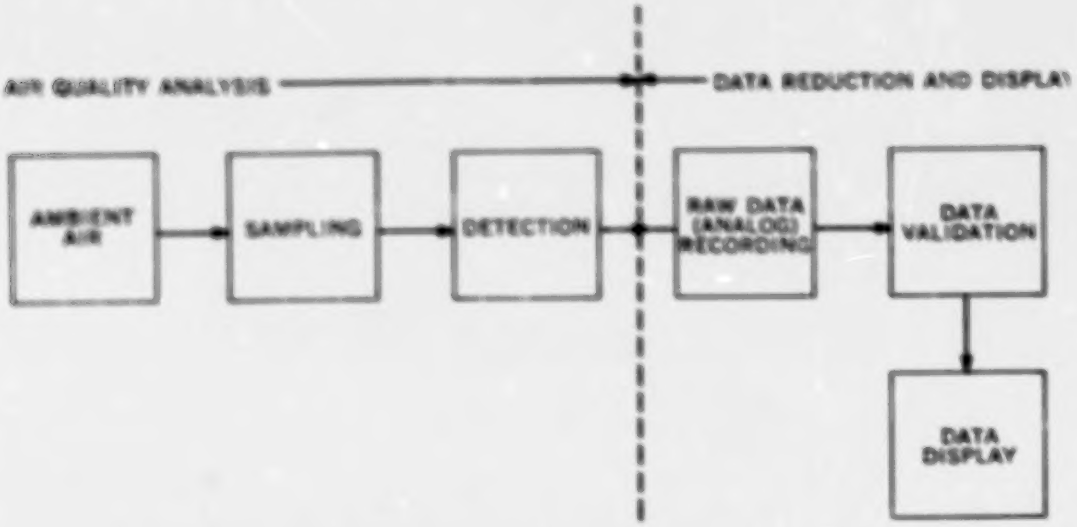


Figure 1 — Air Quality Data Acquisition System

U. S. PUBLIC HEALTH SERVICE CONTINUOUS AIR MONITORING PROGRAM

Objectives of the USPHS Continuous Air Monitoring Program[1] may be stated as follows:

1. To provide information on the concentrations in major American cities of various gaseous air pollutants, which may be related to auto exhaust.
2. To provide continuous data as basic information for research studies, including a study of programming data generation to optimize routine monitoring procedures.
3. To provide basic data for the prediction of dosage levels to which people may be exposed and to which health effects may be related from epidemiological findings.

Six cities (Chicago, Cincinnati, Philadelphia, San Francisco, St. Louis, and Washington, D.C.) were selected by the Public Health Service to provide data directly for these specified objectives. Corresponding data from Los Angeles and Detroit are available from measurements by the Los Angeles Air Pollution Control District and from a study of health effects on animals exposed to urban air by Wayne State University in Detroit.

SAMPLING

Except for minor differences, the stations in five of the cities are essentially the same in that the shelters were built specifically to house the instruments. In the remaining cities facilities already available are utilized. The constructed buildings provide approximately 400 square feet of floor space, part of which is used to accommodate a desk for the technician, who is in daily attendance. Two of the buildings are made of Armco prefabricated metal units and are rectangular in plan. Three of the stations are of Pease geodesic dome construction (Figure 2). Air conditioning and heating are provided, as well as facilities for water, electricity, and sewage disposal. Ambient air is sampled within 10 to 15 feet above ground level at each station through an air intake on top of the building.

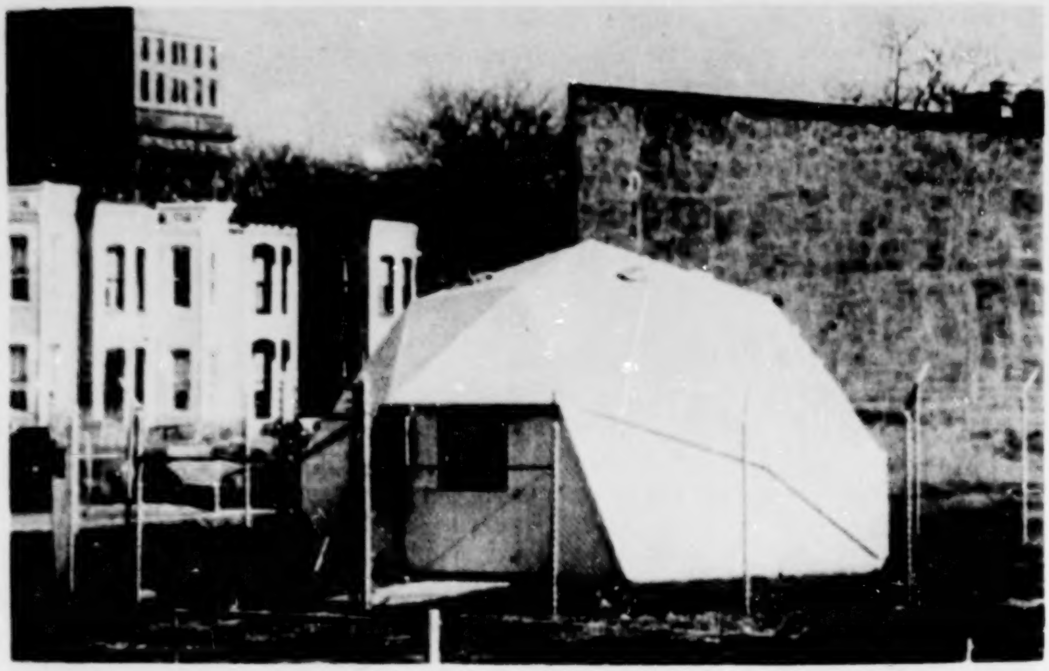


Figure 2 — Pease Dome Air Monitoring Station, Philadelphia, Pa.

These stations are located in downtown areas as close as possible to the center of each city's business district (Table 1). Some of the considerations in selecting a

Table 1 — Sampling Station Sites

City	Building	Location
San Francisco	Garage	Union Square Garage, Inc. Union Square
Chicago	Armco	445 South Plymouth Court
Cincinnati	Armco	Ann and Central Avenues
Philadelphia	Pease	c/o Franklin Institute 2031 Race Street
St. Louis	Pease	215 South 12th Street
Washington, D. C.	Pease	1027 First Street, N.W.
Los Angeles	Laboratory	434 South San Pedro Street (13)
Detroit	Laboratory	St. Antoine and Gratiot

suitable area were openness of the surroundings, availability of utilities, proximity to atypical sources, and approval of city building commission or other authorities. The main criterion was that the air being sampled in these locations is typical and representative of the air to which people are exposed in downtown areas.

Sampling probes, made of unbreakable glass pipe, are used to introduce the ambient air through the center of the roof of the building. Inside the building the 1.5-inch-diameter probe branches into 1-inch-diameter arms, which serve as manifolds from which individual instruments sample (Figure 3).

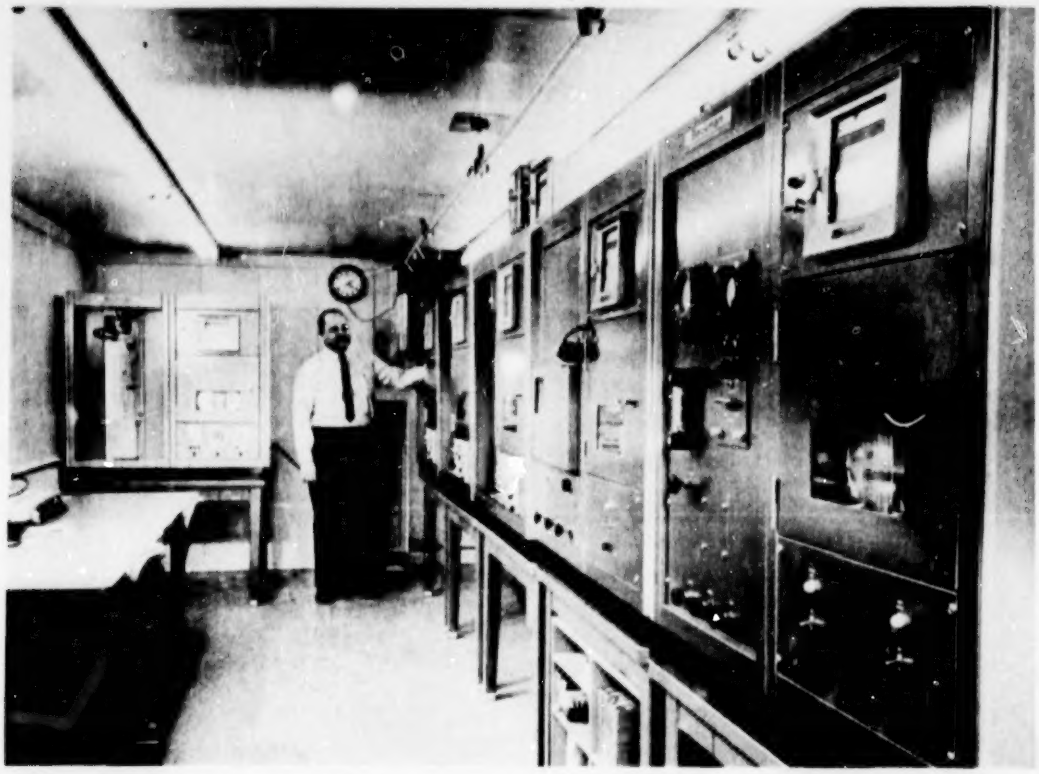


Figure 3 — Interior of the Cincinnati Air Monitoring Station

DETECTION

To eliminate instrument differences as a varying parameter, identical gas analyzers were selected for the six cities for each of the seven pollutant gases. To provide for optimum performance of the analyzers in terms of reliability, sensitivity, stability, etc., a complete set of specifications was written for each type of gas analyzer. Wherever possible, specifications were the same for comparable components of different analyzers to allow for an interchange of components. This uniformity, together with unitized construction in which subassemblies are replaceable as unit components, provides for optimum maintenance and servicing procedures.

The pollutant gases under study are nitrogen dioxide, nitric oxide, sulfur dioxide, total oxidants, carbon monoxide, total hydrocarbons, and ozone. The operation of the analyzers is based on the methods of detection described in the following paragraphs.

Nitrogen Dioxide and Nitric Oxide

The operations of the NO and NO_2 analyzers are interconnected in the sampling operation. Air sampled from the manifold is analyzed for NO_2. The effluent from the NO_2 analyzer is serially analyzed for NO after passing through a potassium permanganate solution (2.5%), which oxidizes the nitric oxide to nitrogen dioxide. Thus, the NO analyzer is essentially an NO_2 detector operating on a pretreated sample.

In both analyzers the NO_2 is reacted with Saltzman reagent to form a visible color.[2] A ratio photometer measures the color change with respect to the unreacted reagent and an electrical analog voltage is generated in a photovoltaic cell. The 90 percent response time is about 15 minutes, the time required for the gas absorption and color formation in the reagent and for the reacted reagent to pass to the point in the analyzer at which the colorimetric detection is made. The concentration range is 0 to 1.0 ppm full scale.

Sulfur Dioxide

Sampled air is passed through a dilute aqueous sulfuric acid solution containing hydrogen peroxide. Absorbed SO_2 is oxidized to sulfuric acid. Concentration of SO_2 is detected as the difference in conductivity in the reagent before and after SO_2 absorption, since the change in conductivity of the solution is proportional to the change in its sulfuric acid content.[3] The conductivity is measured with conductivity cells in a balanced bridge circuit on alternating current. Concentration ranges are 0 to 2 and 0 to 10 ppm full scale. Resonse time for full scale reading is less than 1 minute.

Total Oxidant

Oxidants in the sampled air are absorbed in a buffered 10 percent potassium iodide solution. The reacted solution is measured colorimetrically with respect to fresh reagent by a ratio photometer with filtered light (350-370 mμ).[4] Photovoltaic cells generate the analog voltage. Concentration range is 0.3 ppm midscale and 0.5 ppm full scale. Ninety percent response time is equal to or less than 5 minutes.

Carbon Monoxide

Carbon monoxide analysis is based on the principle of selective absorption of energy by the gas to which the instrument is sensitized.[5] Air is passed through a sample cell, through which infrared energy is transmitted from an ac-powered filament source to a pair of detector cells in series. The detector cells are sensitized with a mixture of carbon monoxide and argon. The sample side of the detector has a lower concentration of CO relative to argon than the reference side. Carbon monoxide in the sampled air is measured as the difference in infrared absorption in the sample and reference detectors. Each detector has a capacitor diaphragm, which moves in response to gas volume changes brought on by infrared energy absorption. Analog voltage is generated in proportion to the difference in energy absorption in the two detector cells, thus giving a measure of CO concentration in the sampled air. Concentration range is 0 to 100 ppm.

Total Hydrocarbons

The operating principle of this analysis is the hydrogen flame ionization technique.[6] Sampled air is mixed with hydrogen and burned in a combustion chamber. Combustion of the hydrocarbon gases in the hydrogen flame increases the production of ions, which are collected at a collector ring near the flame as a result of an electric potential applied between the ring and flame. The migration of ions constitutes an ion current, which is proportional to the carbon atom content of the hydrocarbon pollutant undergoing combustion. Detection of the analog pico amperes generated is by an electrometer. Concentration ranges are 0 to 100 ppm measured as carbon atoms.

Ozone

Sample air is contacted with a solution of potassium iodide to allow a reaction with the ozone pollutant with the liberation of free iodine. About 0.24 volt is applied to a sensor electrode cell, and the polarization current produces a thin layer of hydrogen gas at the cathode. Removal of the hydrogen by its reaction with the free iodine reestablishes the polarization current and the reaction cycle. For every ozone molecule reacting in the sensor, two electrons flow through the external circuit. Thus, electron generation is directly proportional to the oxidant mass concentration. Detection of the electron flow as a function of ozone content in the sampled airflow is by a microammeter. Full-scale concentration is in the range of 0 to 1 ppm. This method of ozone analysis suffers some interference from oxidants such as NO and NO_2. Therefore, it is presently referred to as coulometric oxidants analysis to distinguish it from the colorimetric oxidants analysis.

Response Time

In the analytical methods that involve scrubbing the sampled air in a chemical reagent, there is inherent in the method a minimum amount of integrating or response time such that rapid or peak concentration changes are not resolved but instead are averaged out over the response-time interval. These periods can range from 2 to 15 minutes for the wet chemical methods. For some of the physical methods, such as those for CO and hydrocarbons, the inherent response time is relatively short, about 1 minute.

For the hydrocarbon analyzer a volume container was introduced into the sampling line of the analyzer to give about a 5-minute response time that would correspond to the printout interval. Similar plans are underway for the CO and coulometric analyzers.

The response time for the various analyzers as they are operated in the system is measured from the time the specific pollutant at known concentration is introduced at the sampling probe of the analyzer to the time the analog recorder shows a response equal to 95 percent of the final concentration. The time lag introduced by the sampling manifold is 30 to 45 seconds, which should be added to the following response-time values for the various analyzers:

Analyzer	*Response Time, minutes*
NO, NO_2	15
SO_2	10
Hydrocarbons	5[a]
Oxidants (colorimetric)	5
Oxidants (coulometric)	1
CO	1

[a] Surge bottle attached to sampling line to give longer response time.

Calibration

Calibration is the procedure by which correspondence is established between the electrical analog output of the pollutant analyzer and the pollutant concentration of the air sample entering the instrument.

The broad geographical distribution of the sampling stations and variety of pollutant analyzers within a station necessitated considerable attention to calibration techniques and procedures to assure the collection of accurate and valid data.

The procedure followed in this network involved: (1) the initial calibration, in which a calibration curve was established for each instrument; (2) the standard calibration check, in which periodic checks of the standardized initial calibration are made to allow for drift and variations that may occur in the operation of some of the instrument components over a period of time; and (3) the reference calibration, which is common to all instruments that measure the same pollutant at different stations.

Static calibration techniques are used on the chemico-physical analyzers such as the colorimetric, coulometric, and conductometric instruments. This method of calibration is applied to the detection and recording operations of the data acquisition system, and the sampling and chemical reaction operations are omitted. In one widely used method, a standard solution, chemically equivalent to reagents that have absorbed and reacted with known concentrations of pollutant gases, is substituted in the detection component of the analyzer. The CAMP staff has developed a refinement of this method by using colored pieces of cellophane. These serve as optical filters and reproduce the detected property spectral of optical density. Therefore, these optical filters are checked to determine their pollutant equivalents and are used for static calibration checks on the colorimetric analyzers.

Dynamic calibration applies to all the operations involved in sampling and analysis of the gaseous pollutants. This type of calibration must be made initially in the operation of any instrument.

Availability of gas mixtures containing the desired pollutant gas of known concentration and purity is essential to the dynamic calibration of the CAMP analyzers. The CO and the hydrocarbons analyzers are calibrated dynamically with gas mixtures contained in pressurized cylinders. These gas mixtures are prepared by CAMP personnel or are purchased and analyzed at the Sanitary Engineering Center with an analyzer calibrated against prepared bag mixtures.[10]

Some gas mixtures, such as ozone, cannot be prepared to accurately known concentrations having sufficient stability for dynamic calibrations in the field. In such cases the calibration sample must be analyzed concurrently by an accepted reference method. A dilution board (Figure 4) is used to prepare calibration samples at low concentration for SO_2, NO_x, and O_3. The board provides two sample streams, one for the sampling and analysis instrument under test and one for manual sampling and laboratory analysis, the latter serving as a reference calibration.

DATA RECORDING

All of the various methods of sampling and analysis discussed above generate an electrical analog signal (Figure 5). The analog strip-chart recorders for the various

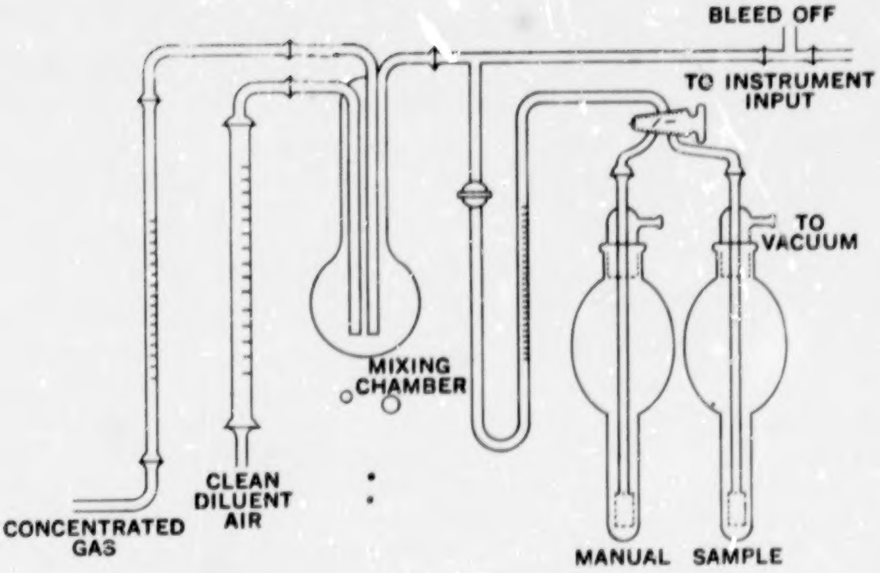

BLEED OFF
TO INSTRUMENT INPUT
TO VACUUM
MIXING CHAMBER
CLEAN DILUENT AIR
CONCENTRATED GAS
MANUAL
SAMPLE

Figure 4 — Dilution Apparatus for Dynamic Referee Calibration

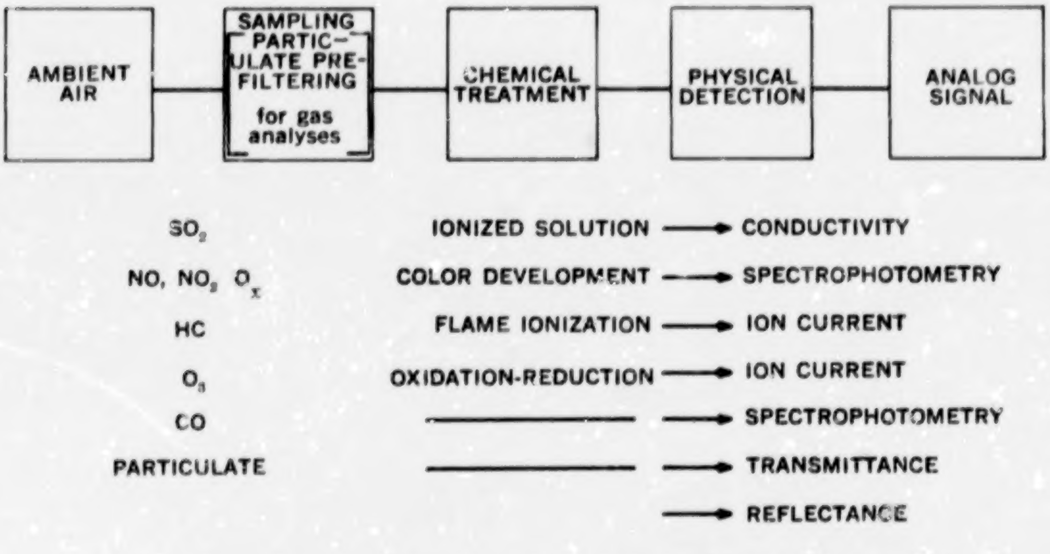

AMBIENT AIR
SAMPLING PARTICULATE PRE-FILTERING for gas analyses
CHEMICAL TREATMENT
PHYSICAL DETECTION
ANALOG SIGNAL
SO2
NO, NO2 Ox
HC
O3
CO
PARTICULATE
IONIZED SOLUTION
COLOR DEVELOPMENT
FLAME IONIZATION
OXIDATION-REDUCTION
CONDUCTIVITY
SPECTROPHOTOMETRY
ION CURRENT
ION CURRENT
SPECTROPHOTOMETRY
TRANSMITTANCE
REFLECTANCE

Figure 5 — Physical and Chemico-Physical Analyzer Systems

gas analyzers, however, are standardized to permit interchangeability and to facilitate servicing and maintenance. The output analog signals of the analyzers are either attenuated or amplified to be compatible with the 0 to 1 millivolt input range of the analog recorders. Thus, in the data handling operations from introduction of the analog signal into the analog recorder to the final data display operation, all operations are common to each of the pollutant gases (Figure 6).

Analytical data obtained in monitoring air pollution are usually presented in the form of continuous strip-chart recordings. This system offers several advantages: (1) a graphic display that can be scanned visually for immediate interpretation of the data; (2) relatively instantaneous values in the form of a continuous record that is easily checked for anomalous data, which might represent malfunction of the detection system; and (3) a fairly reliable and accurate measurement system for a nominal price. Strip-chart recorders were obtained for these analyzers primarily for the advantage offered in item (2).

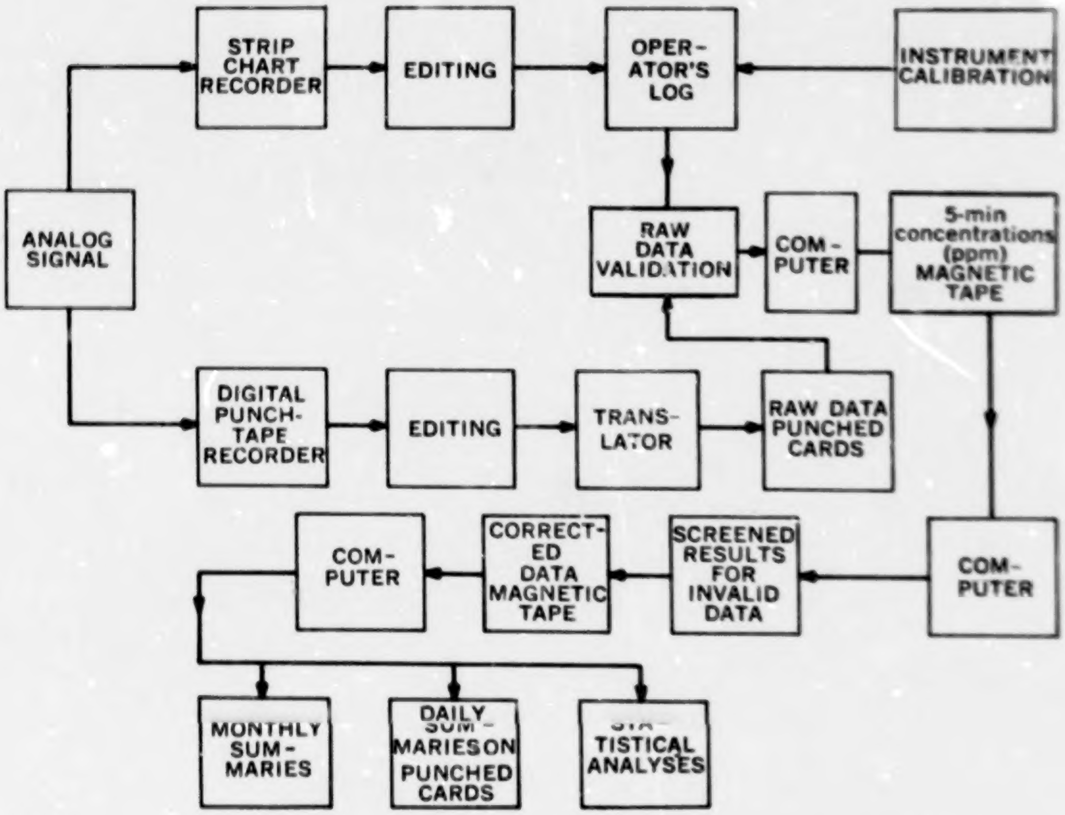

STRIP CHART RECORDER
EDITING
OPERATOR'S LOG
INSTRUMENT CALIBRATION
ANALOG SIGNAL
RAW DATA VALIDATION
COMPUTER
5-min concentrations (ppm) MAGNETIC TAPE
DIGITAL PUNCH-TAPE RECORDER
EDITING
TRANSLATOR
RAW DATA PUNCHED CARDS
COMPUTER
CORRECTED DATA MAGNETIC TAPE
SCREENED RESULTS FOR INVALID DATA
COMPUTER
MONTHLY SUMMARIES
DAILY SUMMARIES ON PUNCHED CARDS
STATISTICAL ANALYSES

Figure 6 — Data Reduction and Display, USPHS Continuous Air Monitoring Program

The tremendous quantities of data being acquired in this project prohibit the use of the normal procedure, which requires manual or semi-automatic reduction from the strip-chart recording of digital data onto punched cards, punched tape, or magnetic tape compatible with the input of electronic computers. For the seven gases continuously monitored in the eight cities every 5 minutes throughout the year, 112,896 items of data are generated in a single week, or approximately 6 million items in a year. To overcome the problem of handling strip-chart data, a digital punch-tape recorder has been incorporated with each strip-chart recorder.

This digital punch-tape recorder is a modification of the Fischer and Porter analog-to-digital recorder (ADR) designed for rotary shaft input (Figure 7). The modification (henceforth referred to as the modified ADR) involves the addition of a servomechanism assembly and related electronic circuitry. A retransmitting slidewire on the strip-chart recorder drives the modified ADR.

The ADR input is an angular positioning of two digitally encoded wheels geared together in a 100:1 ratio. Each of the wheels presents two digits in range from 00 to 99 in one revolution. The wheels are marked so that a visual reading is indicated at all times. An electric timer programs the punch mechanism to punch the digital data directly on paper tape every 5 minutes.

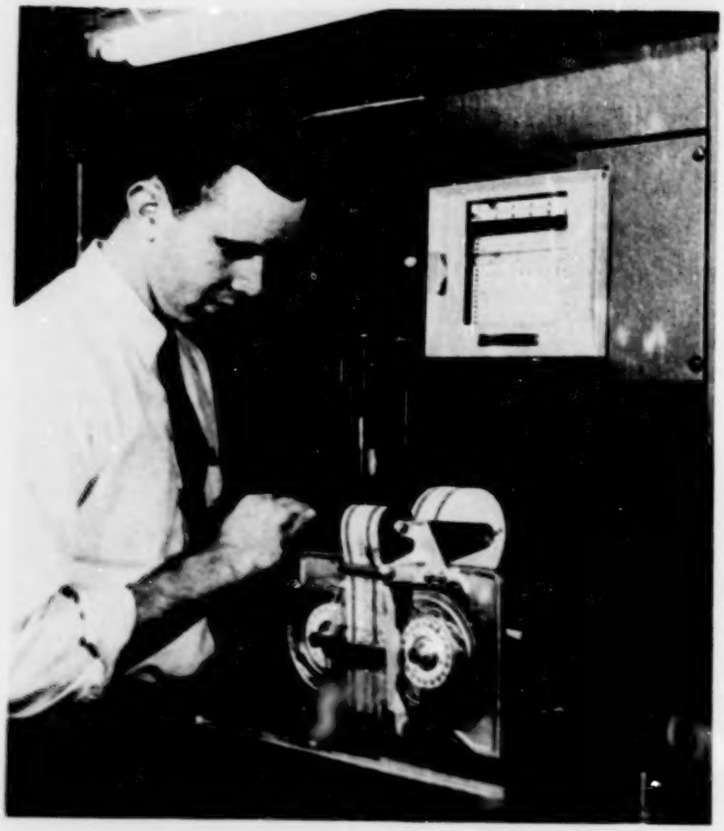


Figure 7 — Analog-to-Digital Punch-Tape Recorder

Paper tape is provided with hourly interval markings and 12 punch spacings to accommodate 12 items of data programmed within the hour. Synchronization of data punching with time of day is checked visually, and any malfunction of equipment is detectable.

The servomechanism modifies the rotary shaft input requirements of the ADR so that analog information existing as ac voltage on the retransmitting slidewire can be converted directly to digital data on punched tape.[9] In principle this system utilizes the null-balancing-type circuit, consisting of a voltage amplifier (amplifier of a strip-chart recorder is automatically switched in to serve in this capacity) and a balancing potentiometer and motor, both coupled to the input shaft of the ADR. The modified ADR is designed to give 3-digit full-scale output (000 to 999) for full-scale signals of 1.0 millivolts.

For a changing signal level at the input, the servomechanism will continuously follow to maintain a null balance and consequently cause the ADR to give the correct instantaneous output at all times. The digital tape punch of the analog input takes place only when the programmer commands the ADR to punch a reading on the paper tape, at which time the encoder wheels lock in place and the reading at that instant is punched.

DATA VALIDATION

The digital punch-tape recordings are forwarded weekly from each station to the Robert A. Taft Sanitary Engineering Center for evaluation. These data are uncorrected upon arrival and as such, coming directly from the analyzer, are treated as "raw" data. The technician at each station maintains a daily operator's log on each instrument for each gas pollutant. This log includes such things as calibration checks, zero-drift corrections, instrument malfunctions, and bad data recording as indicated by the strip-chart recorder. An operator's log sheet and a strip-chart record accompanies each corresponding punch-tape record sent to the Center.

Punch-tape data received at the Center are transferred directly onto IBM cards by means of the Fischer and Porter translator so that "raw" data are in a form compatible with electronic computers. The translation from tape to cards is achieved by three component instruments: a programmer and reader, which together comprise the translator, and an IBM key-punch machine. The programmer provides the time and date the data were taken and a 7-digit identification code, 5 digits to identify the city and 2 digits to identify the pollutant. The sequencing of time and date is automatically programmed. The reader transfers the punch-tape data of gas concentrations to the key-punch unit, in which 1-hour sets of readings per gas (12 items of data) are tabulated per card. The resulting deck of IBM cards incorporates the "raw" data in a form that can be handled directly by a computer. The "raw" data together with the calibration information and corrections indicated in the operator's log are programmed through the computer to give raw data on 5-minute concentration values in ppm on magnetic tape. These concentration values are subsequently screened by a computer program for invalid data to give corrected data on magnetic tape.

DATA DISPLAY

Computer programs have been prepared to present the corrected data in the form of monthly summaries and daily summaries on punched cards, and for various statistical analysis studies such as dosage, exposure time, frequency distribution, etc.

LOS ANGELES COUNTY APCD STATION NETWORK

Since its inception the Los Angeles County Air Pollution Control District has employed an air monitoring network to ascertain the magnitude and character of the air contaminants. This network has varied both in number of stations and scope of sampling.[11] Late in 1961 a comprehensive physiological study of the problem of air pollution from auto exhaust was started by the University of Southern California under a contract with the U.S. Public Health Service. Four locations in the Los Angeles Basin were selected to conduct the study, in which experimental animals were to be exposed to the sampled air. A separate but corollary contract was executed between LACAPCD and USPHS to provide three of the exposure sites and to provide air monitoring services at all locations.[12] During 1962 three of the LACAPCD Network Stations were relocated to meet the requirements of the USC-PHS Study.

LACAPCD currently operates 10 air monitoring stations within the confines of the County. The objectives of the continuously recording automatic instrument systems employed in this program may be summarized as follows:

1. To support research efforts such as trend evaluation and meteorological, atmospheric chemistry, and animal exposure studies.
2. To implement an emergency regulation pertaining to the buildup of certain known toxicants.
3. To ascertain the effectiveness of air pollution control regulations.

SAMPLING

Currently, LACAPCD Network provides coverage of the Los Angeles Basin for monitoring air quality by means of 10 stations (Table 2), which include the four sites selected for the physiology study. Each of the four sites was selected to fulfill a requirement of the study: the USC Medical School site to represent high levels of both primary and photochemical automotive-related air pollution; the Burbank site to represent a fairly densely populated suburban area subject to relatively high levels of air contaminants; the Azusa site to represent a pollution receptor area as designated by previous air quality measurements and further defined as one in which measurements of photochemical pollutants such as ozone are relatively high while those for CO and NO_x are relatively low; and a freeway to represent a [illegible]tion adjacent to a major traffic artery.

Table 2 — Air Monitoring Installations

No.	Location	Street Address	Telephone
1	Downtown Los Angeles	434 South San Pedro Street	MAdison 9-4711 Ext. 66032
51	El Segundo	359 Maryland Street	
60	Azusa	803 Loren	CUmberland 3-5967
64	Pasadena	862 East Villa	MUrray 1-8748
68	Inglewood	5037 West Imperial Highway	ORegon 8-6362
69	Burbank	228 West Palm Avenue	VIctoria 9-3642
70	General Hospital	1411 North Eastlake Avenue	225-4085
71	West Los Angeles	2351 Westwood Boulevard	478-6754
72	Long Beach	3648 Long Beach Boulevard	424-5420
	Freeway Site	608 Heliotrope Drive	666-2672
	CONTROL CENTER	434 South San Pedro Street	MAdison 9-4711 Ext. 66011

A pyrex-pipe manifold system is used to introduce sampled air to the analyzers. Air is sampled both from the outside ambient air and from the purified-air control rooms in the animal studies. In the latter case, for a relatively continuous check of the contaminant level, a sampling valve operated by a timing circuit obtains samples alternately from the ambient air and from the control room. Cycling time permits calculation of hourly averages for each sample. This sampling technique is limited to analyzers having relatively short response time, such as the hydrocarbon instrument. The longer response time of the oxides of nitrogen instruments requires grab sampling in the control room and subsequent laboratory analyses.

DETECTION

For the greater part, the LACAPCD Network analyzers are similar in operating principles to those discussed for the USPHS CAMP analyzers. The flame ionization hydrocarbon analyzer, coulometric ozone, colorimetric oxidant, conductometric SO_2, and colorimetric oxides of nitrogen instruments are the same type as the CAMP instruments. The remaining instruments have been described elsewhere in some detail,[12, 13] and only essential differences from USPHS CAMP equipment will be mentioned here.

Ozone

Measurement of ozone by the ozone photometer is based on ultra-violet absorption by the ozone at wavelength of 2537 angstroms. A dual-cell differential detector measures the ambient air stream against a parallel air stream from which the ozone has been removed by catalytic decomposition by means of a manganese dioxide coated tube. The difference in UV absorption is a linear function of the ozone concentration.

Carbon Monoxide

Carbon monoxide is measured in a nondispersive infrared analyzer. This analyzer incorporates a parallel pair of detector cells in a dual-beam arrangement as contrasted with the series detector cells and single beam of the analyzer in the CAMP instrument.

Particulate

An automated particulate sampler and analyzer determines filterable black aerosols by light reflectance and transmittance immediately after the sample is collected. Air is sampled at 25 cfh through a paper filter medium, which is advanced intermittently. Since the flow rate changes slightly during the sampling interval, three flow rate measurements are made during each sampling period of 1 hour and the average flow rate calculated. Detection is by means of a photovoltaic cell, which generates an analog voltage signal as a function of the light reflected from or transmitted through the filtered sample spot of particulates.

Calibration

The chemico-physical systems, i.e., oxidants, ozone, oxides of nitrogen, and sulfur dioxide are calibrated dynamically by use of a dilution system, and samples are obtained simultaneously for subsequent reference analysis as discussed previously. CO and hydrocarbon analyzers are calibrated dynamically by use of prepared known gas mixtures in pressurized cylinders.

DATA RECORDING AND VALIDATION

Analog data generated in the Los Angeles Network are recorded on strip-chart recorders. Only the Downtown Station has been equipped with the ADR recorders to provide digital data directly. For the remaining nine stations the analog data are edited manually and, with supplemental field reports, digital tabulations are prepared manually for a key punch to produce corrected and validated data (Figure 8).

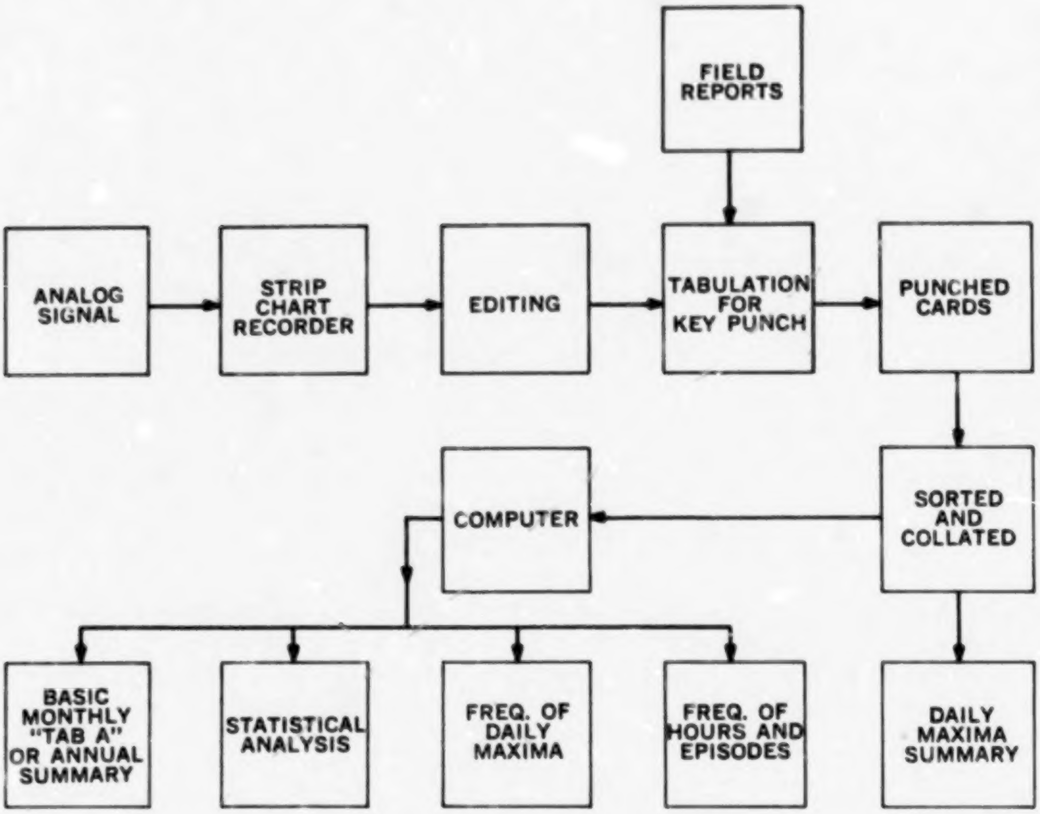

FIELD REPORTS
ANALOG SIGNAL
STRIP CHART RECORDER
EDITING
TABULATION FOR KEY PUNCH
PUNCHED CARDS
COMPUTER
SORTED AND COLLATED
BASIC MONTHLY "TAB A" OR ANNUAL SUMMARY
STATISTICAL ANALYSIS
FREQ. OF DAILY MAXIMA
FREQ. OF HOURS AND EPISODES
DAILY MAXIMA SUMMARY

Figure 8 — Data Reduction and Display, Los Angeles County Air Pollution Control Network

DATA DISPLAY

From the corrected data on punched cards, a summary of daily maxima is obtained directly after sorting and collating. A computer program applied to the collated punch cards provides Basic Monthly "Tab A," Statistical Analyses, frequency of daily maxima, and frequency of hours and episodes.

CALIFORNIA STATE HEALTH DEPARTMENT

A brief review of early plans of the California State Health Department will be of interest with respect to design of their data handling system. Five air pollutants will be recorded: carbon monoxide, hydrocarbons, oxides of nitrogen, and oxidant. The data-generating analyzers are very much the same as those in the Los Angeles Network; these are located in a number of stations (about 15) throughout California. A station at Oakland is now being equipped with digital punched-tape apparatus as a pilot study. Tentative plans are that the digital raw data will go through a computer program in conjunction with operator's information to give pollutant concentrations on punched cards. These data will be stored by a computer on magnetic tape and will be available for computer programs to give various summaries (Figure 9).

The data-logging system consists of Coleman Digitizers attached to the analog strip-chart recorders; a Coleman Data Processor, which samples the Digitizer shaft positions in sequence; a Coleman Tape-Punch Control Unit, which encodes data from the Data Processor and data from a digital calendar and from a digital clock for entry into a Friden motorized tape punch.

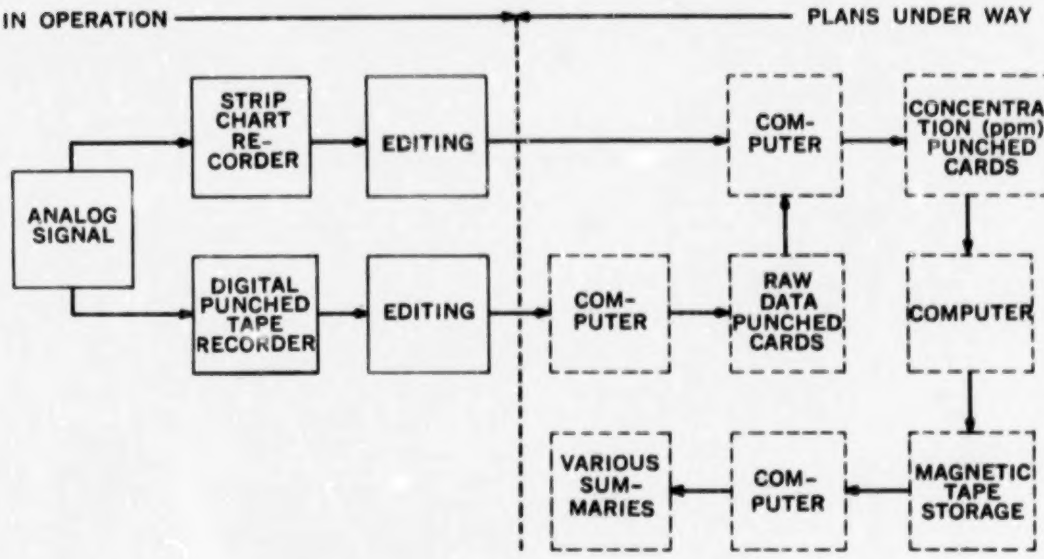

IN OPERATION
PLANS UNDER WAY
ANALOG SIGNAL
STRIP CHART RE-CORDER
EDITING
COM-PUTER
CONCENTRATION (ppm) PUNCHED CARDS
DIGITAL PUNCHED TAPE RECORDER
EDITING
COM-PUTER
RAW DATA PUNCHED CARDS
COMPUTER
VARIOUS SUM-MARIES
COM-PUTER
MAGNETIC TAPE STORAGE

Figure 9 — Data Reduction and Display, California State Health Department

The eight-channel punch tape is standard and operates with serial entry as contrasted with the ADR sixteen-channel punch tape with parallel entry. Five channels are used for binary coded digits: 0,1,2,4,8. One channel is for parity check and another for "end of line" to identify the end of the cycle. The remaining channel is unwed.

The tape entry for one cycle has the following sequence (Figure 10):

1. Station number	— two digits
2. Date and time	— eight digits
3. Pollutant identification	— one digit
4. Mode of operation	— one digit
5. Pollutant concentration	— three digits

Items 3, 4, and 5 are repeated for each of five pollutants within the cycle.

6. End of cycle identification — one digit

This gives a total of 36 character words per cycle. A complete cycle is punched out within 6 seconds exclusive of any balancing time delay added by the strip-chart recorders. The digital recording of the five pollutant readings is made at 5-minute intervals.

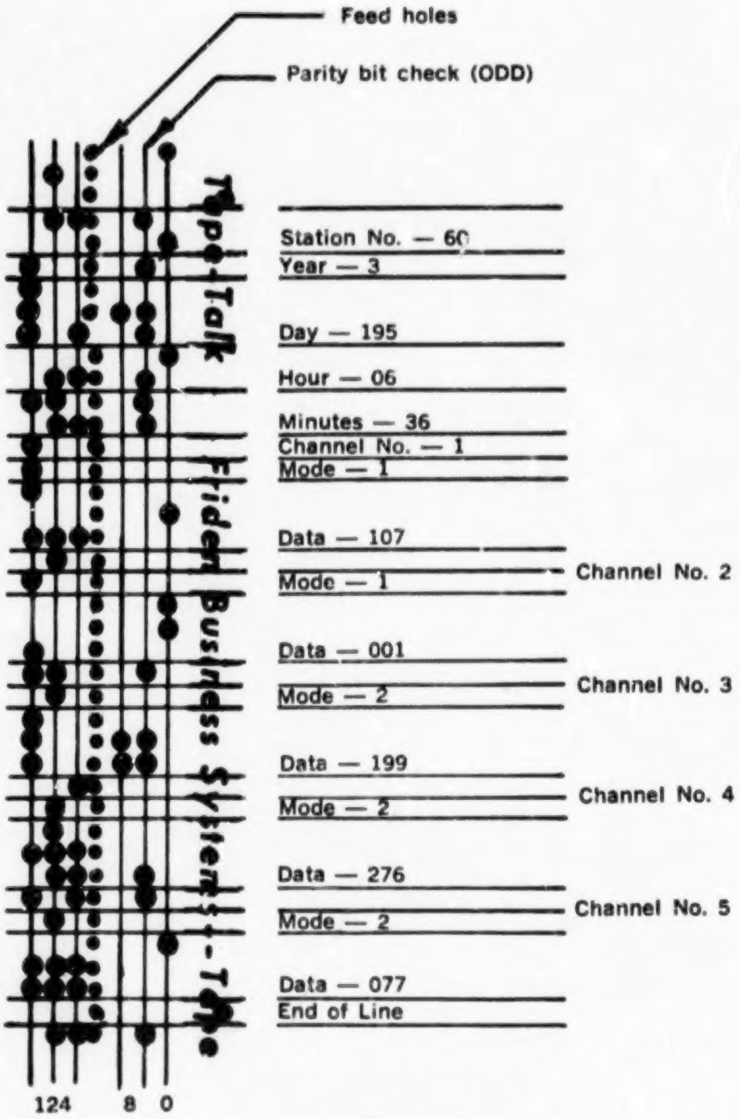

Feed holes
Parity bit check (ODD)
Tape-Talk | | Friden Business Systems - Tape
Station No. — 60
Year — 3
Day — 195
Hour — 06
Minutes — 36
Channel No. — 1
Mode — 1
Data — 107
Mode — 1
Channel No. 2
Data — 001
Mode — 2
Channel No. 3
Data — 199
Mode — 2
Channel No. 4
Data — 276
Mode — 2
Channel No. 5
Data — 077
End of Line
124
8
0

Figure 10 — Sample Tape From Friden Tape Punch

REFERENCES

1. Jutze, G. A. and Tabor, E. C. The Continuous Air Monitoring Program. JAPCA. 13, 6:278. 1963.

2. Saltzman, B. E. Colorimetric Microdetermination of Nitrogen Dioxide in the Atmosphere. Anal. Chem. 26, 1949. 1954.

3. Thomas, M. D. The Sulfur Dioxide Autometer. Proc. of Instr. Soc. of America. Vol. II, Part I. 1956.

4. Littman, F. E. and Benioliel, R. W. Continuous Oxidant Recorder. Anal. Chem. 25, 1480. 1953.

5. Waters, J. L. and Hartz, N. W. An Improved Luft-Type Infrared Gas and Liquid Analyzer. Instrument Society of America Meeting, Houston. 1951.

6. Morris, R. A. and Chapman, R. L. Flame Ionization Hydrogen Analyzer. JAPCA. 11, 467. 1961.

7. Mast, G. M. A New Ozone Meter. Summer Instrument and Automation Conference of the Instrument Society of America, San Francisco. 1960.

8. Christman, K. F. and Foster, K. E. Calibration of Automatic Analyzers in a Continuous Air Monitoring Program. Presented at the Air Pollution Control Association Annual Meeting, Detroit, Mich. 1963.

9. Nader, J. S. and Coffey, W. L. Direct Digital Recording of Air Pollution Measurements. Presented at Air Pollution Control Association Annual Meeting, New York City. 1961.

10. Altshuller, A. P., et al. Storage of Vapors and Gases in Plastic Bags. Intern. J. Air and Water Pollution. 6, 75. 1962.

11. Taylor, J. R. Methods and Procedures Employed in the Recordation and Processing of Air Quality Data. Analysis Paper No. 35, LACAPCD, Los Angeles, California. August 10, 1960.

12. Bryan, R. J. Instrumentation for an Ambient Air, Animal Exposure Project. JAPCA. 13, 6:254. 1963.

13. Bryan, R. J. and Romanovsky, J. C. Instrumentation for Air Pollution. Instruments and Automation. 29, No. 2, December 1956.

Dr. Harrison E. Cramer
Director, Round Hill Field Station
Massachusetts Institute of Technology, Cambridge

SUMMARY

Automatic data collection and data processing techniques have found in the past decade important application in empirical studies of low-level atmospheric structure and in diffusion problems associated with the operation of nuclear reactors. A typical data acquisition system comprises four major subsytems: sensors, telemetry, central control, and displays. A review of basic features of existing acquisition systems at several installations shows a wide variety of subsystem designs and focuses attention on factors that must be considered in the selection of system components: sensor response characteristics, sensor location and density, data sampling rates, parity checks, and time or space averaging techniques. Acquisition systems designed for use in air pollution studies or control should be capable of handling three scales of meteorological information: macroscale, mesoscale, and microscale. To illustrate the application of engineering design criteria, an idealized system is described in detail.

DATA ACQUISITION SYSTEMS IN METEOROLOGY

INTRODUCTION

In meteorological studies of atmospheric diffusion mechanisms and the structure of turbulence, automatic data collection and data handling techniques have become practically indispensable. The relatively large number of observations required for a statistically significant description of characteristic air properties effectively precludes the use of manual techniques for data acquisition, reluction, and analysis. This trend has been facilitated by the vastly increased capability of small computers that have been introduced in the last 3 or 4 years. The introduction of automatic techniques in meteorological instrumentation systems has proceeded rather cautiously, in part because of economic factors. Piecemeal procedures and general lack of over-all planning and system engineering characterized many data acquisition system developments in the past. Basic uncertainties as to the specific operational or research requirements to be met by the measurement system and the time and space variability of meteorological parameters have contributed to this situation.

A typical data acquisition system comprises four major subsystems: *sensors*, which provide electrical or mechanical analogs of meteorological variables; *telemetry*, which provides for the transfer of sensor information to a central collection point; *central control*, which provides for the interrogation of sensors, the recording and processing of sensor outputs, and the routing of processed data to displays or storage; and *displays*, which present processed data in a convenient form. Figure 1 illustrates these relationships. This paper begins with a brief summary of the design and operation of existing systems at various locations. Next, some of the fundamental problems associated with the measurement of meteorological variables and the basic operations involved in a data acquisition system are described. Finally, the application of engineering analysis techniques to system design is illustrated by consideration of a hypothetical system designed to serve the need of an urban air pollution study.

SURVEY OF EXISTING DATA ACQUISITION SYSTEMS

One of the first automatic data acquisition systems was installed at Dugway Proving

Ground approximately 10 years ago, primarily for the purpose of collecting information useful in small-scale climatological studies. In this system, measurement of wind speed, wind direction, air temperature, vertical temperature gradient, surface pressure, relative humidity, and radiation at a number of widely separated stations were periodically transmitted in a digitally coded form over telephone lines to a central collection point. Here the data were decoded and printed out as numerical sequences by an electric typewriter. One of the main difficulties experienced in the use of this system was the lack of an adequate facility for translation of the acquired data into a record-form that can be processed by an automatic computer.

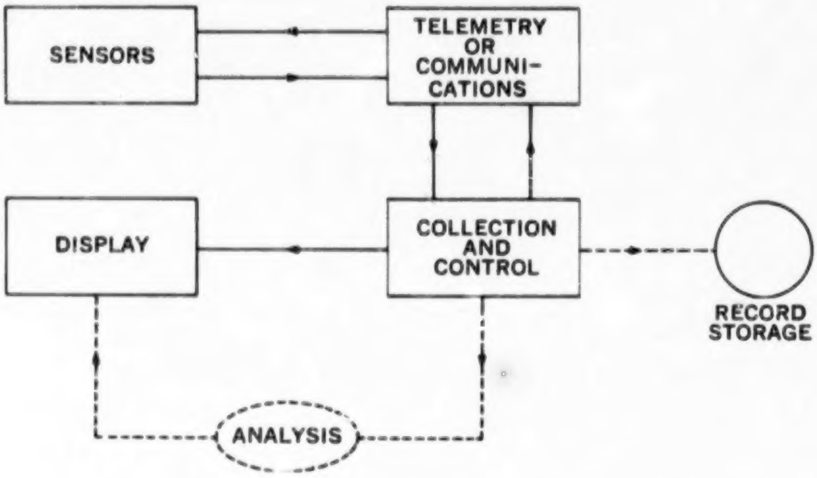


Figure 1 — Schematic Diagram of Basic Components for Data Acquisition System

During the past 10 years other acquisition systems have been developed at Brookhaven National Laboratory (Brown, 1959), Round Hill Field Station (Cramer, Record, Tillman and Vaughan, 1961), Argonne National Laboratory (Moses and Kulhanek, 1962), Oak Ridge, Tennessee (Meyers, 1956), National Reactor Testing Station, Idaho (Islitzer) and other places. These systems have all been aimed at producing a punched paper tape suitable for direct processing by an automatic digital computer. The Argonne Laboratory system contains an automatic programmer, which sequences through readings of the various meteorological sensors on a preset schedule controlled by a digital clock. This system punches a paper tape, which is read directly through a teletype tape reader and printed. The tape is subsequently converted to punched cards for processing by an automatic computer. Characteristics of some of these systems are illustrated in Figure 2 and Table 1.

The first serious step toward actual on-line (*real-time*) computation of data was the Air Force WIND system, which became operational in 1961. This computer-controlled system automatically acquires micrometeorological data and provides diffusion-prediction information for operational use on a continuously updated basis (Haugen, Meyers, Taylor, 1962). Information from the various meteorological sensors is transmitted in analog form over wire lines to an analog-digital converter controlled by the computer. The sensor multiplexing, or switching, is controlled from the computer; all readings are directly processed in the computer, which performs the necessary diffusion computations and punches summary data on a teletype tape and a typewriter. Although, from a modern system engineering point of view many of the components of the WIND sys-

Table 1 — Idaho Falls Data Acquisition System (Fast System)

Parameter	Integration Period	Sampling Period, Minutes	Accuracy
Temperature	Instantaneous	10 Min	1.0°F
Temperature Gradient	Instantaneous	10 Min	0.1°F
	Also 60-Min Avg	60 Min	
Solar Radiation	60-Min Avg	60 Min	3%
Dew Point	Instantaneous	10 Min	1.0°F
Wind Speed	10-Min Avg	10 Min	3%
	60-Min Avg	60 Min	
Wind Direction	10-Min Avg	10 Min	±3.5%
	60-Min Avg	60 Min	
Vertical Wind Direction	Continuous	Continuous	1.0°
	2-Min Avg	2 Min	
	10-Min Avg	10 Min	
	30-Min Avg	30 Min	
	60-Min Avg	60 Min	
	Manual	Manual	

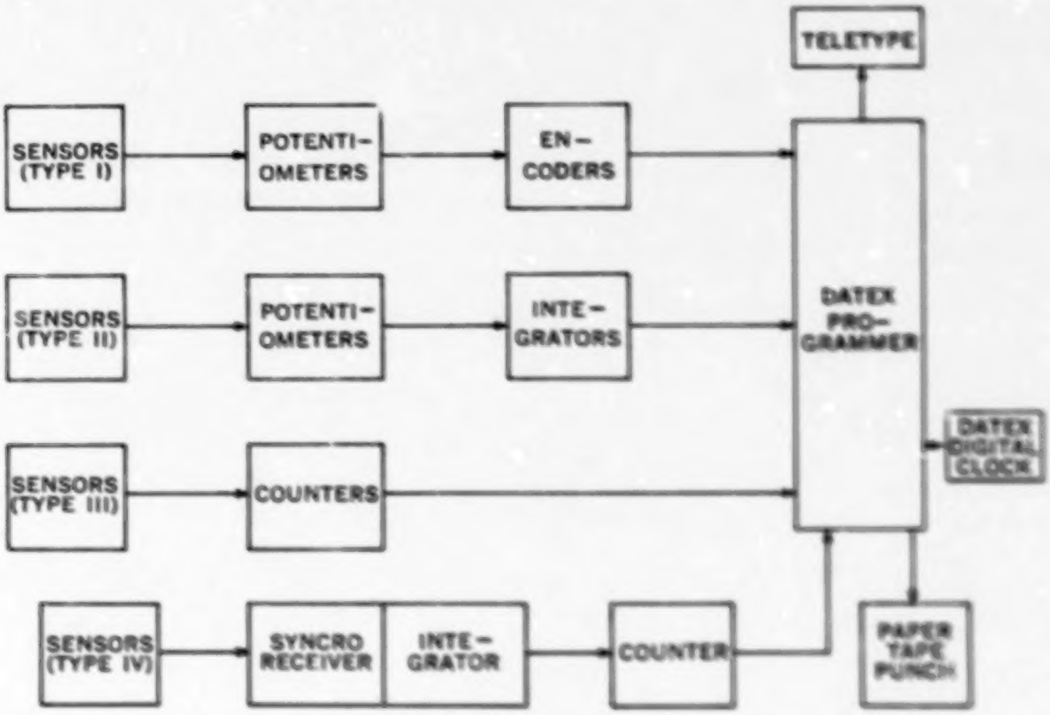

TELETYPE
SENSORS (TYPE I)
POTENTI-OMETERS
EN-CODERS
SENSORS (TYPE II)
POTENTI-OMETERS
INTE-GRATORS
DATEX PRO-GRAMMER
DATEX DIGITAL CLOCK
SENSORS (TYPE III)
COUNTERS
SENSORS (TYPE IV)
SYNCRO RECEIVER
INTE-GRATOR
COUNTER
PAPER TAPE PUNCH

Figure 2 — Simplified Block Diagram Argonne Meteorological Data Processing System

tem are far from up to date, it is giving faithful service at both Cape Kennedy and Vandenberg installations and represents a significant step in philosophy and approach to the handling of this type of meteorological data. A simplified schema of the data flow in WIND is shown in Figure 3.

There are currently underdevelopment at various national test ranges computer-controlled meteorological data collection systems of substantially greater capacity than the systems now existing, both in terms of numbers of sensors under control and the rate at which these sensors are read. The current trend is strongly toward on-line computer control of these functions. The advent of the digital computer as a central component in electronic systems is a phenomenon of the past 5 to 10 years that is only now beginning to be fully appreciated in the area of meteorological instrumentation. Flexibility of the modern digital computer enables it to replace literally hundreds of the special-purpose devices formerly used for the acquisition, filtering, transmission, recording, and processing of data. The replacement of numerous small components and the elimination of the interface problems, which proliferate when large quantities of electronic gear are tied together, will usually more than offset the expense of the computer. At the same time the increase in automation makes it incumbent upon the user or the system planner to specify more carefully, in advance of equipment purchases, the system functions and the engineering philosophy to be followed.

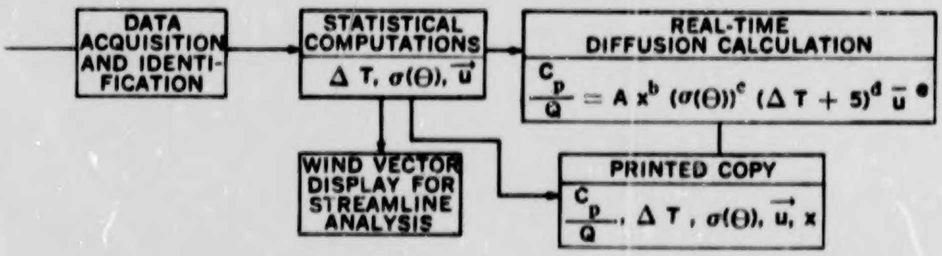

DATA ACQUISITION AND IDENTIFICATION
STATISTICAL COMPUTATIONS
$\Delta T, \sigma(\Theta), \vec{u}$
REAL-TIME DIFFUSION CALCULATION
$\frac{C_p}{Q} = A x^b (\sigma(\Theta))^c (\Delta T + 5)^d \bar{u}^e$
WIND VECTOR DISPLAY FOR STREAMLINE ANALYSIS
PRINTED COPY
$\frac{C_p}{Q}, \Delta T, \sigma(\Theta), \vec{u}, x$

Figure 3 — Simplified Block Diagram Real Time Data Processing for WIND

BASIC CONSIDERATIONS IN THE DESIGN AND OPERATION OF METEOROLOGICAL DATA ACQUISITION SYSTEMS

GEOPHYSICAL CONSIDERATIONS

The basic purpose to be served by a meteorological data acquisition system is generally to provide a satisfactory description of atmospheric structure within a specified reference volume. In air pollution problems the horizontal dimensions of the reference volume are fixed by the areal extent of an urban complex, for example, and the vertical dimension is set by the maximum height attained by any pollutants that are transported across the complex by the wind. Relevant properties of atmospheric structure within the reference volume include the three-dimensional distribution of mean air temperature, moisture, wind speed, wind direction, and the turbulent fluctuations of the latter two variables. In mathematical terms, the mean value, $\overline{M}$ of a meteorological variable obtained from a time series at a fixed point is expressed as

$$\overline{M} = \frac{1}{T} \int_{t_o - T/2}^{t_o + T/2} M(t)\,dt \qquad (1)$$

where T is the length of record and t_o is an arbitrary reference. It also follows from the argument presented above that the value of M at any arbitrary time t is given by

$$M = \overline{M} - M' \qquad (2)$$

and

$$\overline{M'} = 0 \qquad (3)$$

where M' is the departure of M from the mean. Generally, M is a function of both space and time variables:

$$M = f(x,y,z,t) \tag{4}$$

The choice of appropriate integration limits is dictated both by the scale of the problem to be investigated and the form of the energy spectrum of the meteorological variables. Most characteristic air properties exhibit a spectrum of variability that is at least quasi-continuous (consists of disconnected, continuous segments) over a very wide range of time or space frequencies. Since the spectrum is quasi-continuous over a broad range of frequencies, measurements of its properties is limited at high frequencies by the response characteristics of the instrumentation and at low frequencies by the length of record. Pasquill (1962) has shown that the omitted portions of the spectrum can be approximated by weighting functions of the form

$$\frac{\sin^2 \pi n t}{(\pi n \bar{t})^2} \quad \text{and} \quad \frac{1 - \sin^2 \pi n \bar{T}}{(\pi n \bar{T})^2}$$

where $\bar{t}$ is the response time of the measurement system, $\bar{T}$ is the length of record, and n is an arbitrary frequency.

To derive meaningful relationships between meteorological variables and diffusion patterns, for example, it is usually necessary to choose an averaging time $\bar{T}$ such that $\bar{M}$ is reasonably stable.

The space and time variability of meteorological variables is only partially understood at present, and such choices as we have been discussing are usually not routine. Determination of upper and lower frequency limits that will include the "significant" portion of the spectrum for a given variable is still, in the last analysis, a matter for experienced judgment.

Because of the importance of these scaling considerations, it is usually necessary in meteorological applications to relate the functional requirements to three scales of observations: macroscale, mesoscale, and microscale. Macroscale data are those normally used in describing the general weather conditions prevailing over a large area, perhaps 100 miles on a side. Mesoscale data pertain to the general environmental conditions within a few miles distance and in particular to the deviations of the local parameters from the general macroscale weather. Microscale data deal primarily with the fine structure of the local atmosphere for distances of 1 mile or less. With each of these scales of observation are associated different sensor and system input requirements, different data rates, and different processing and display requirements. A properly conceived system will provide for adequate integration of all three data streams.

SYSTEMS ENGINEERING CONSIDERATIONS

The proper development of an automatic data acquisition system for meteorological use involves both meteorology and data systems engineering techniques. Failure to recognize this fact at the outset results in a system based on many practial compromises that may fail to meet the application requirements optimally or even at all. The engineering design of such a system should reflect both the realities of current data systems technology and the ultimate application for which the measurement system is intended. It is well worth the effort to develop a systematic plan for the implementation of the system in advance of choosing specific pieces of hardware. This effort requires technical skills in the areas of operations research, communications systems, computer systems engineering, and programming as well as meteorology to develop a definition of the system concept adequate for determination of subsystem specifications and requirements.

The first step in such a study involves ascertaining and specifying precisely what is to be measured, when, where, and why. In particular, this includes descriptions of the significant range of the spectrum of each variable, the appropriate averaging times, and the portion of the complete time cycle during which information on each variable is wanted. On this basis one can determine data rates in the various parts of the information-gathering network — the fundamental consideration upon which engineering design considerations must be based. It is, of course, also the basis for specification of the meteorological sensors that comprise the sensor subsystem of the data system. The next step involves the specification of the engineering philosophy and logical procedures to be followed in data transmission, as dictated by the inherent data rates; by requirements for mobility, expandability, and change in the system; and by cost. For example, it makes considerable difference whether each meteorological measurement is to be recorded on a punched paper tape for later computer analysis and study, or whether these measurements are connected on-line to a computer for automatic real-time sorting and processing. The choice should be based upon a broad, scientific examination of the total information-handling problem and total cost. Frequently it is less expensive to replace a variety of special devices with a central general-purpose control element, even though the speed and flexibility requirements of the system do not require this.

The third step in this activity entails analysis of data recording and data processing requirements, including necessary mathematical modeling and calculations. A fourth step involves a thorough analysis of requirements for display of the processed information to a potential user.

Such information, properly organized to reveal the interdependence of the various functions, in proper detail, is the necessary foundation for specifying system performance criteria; it is also essential for an organized approach to the technical problems of subsystem hardware and for the integration of subsystem interface requirements into the design approach. It is of the utmost importance for the success of a given meteorological system that this be done in advance of procuring pieces of equipment for the system and in the light of the broader uses of the meteorological measurement program that this system is intended to implement.

The foregoing factors must be investigated against the background of various constraints on the design and implementation of the system. These constraints reflect known limitations in physical, engineering, economic, and human factors, which will significantly affect the feasibility of utilizing specific techniques or equipment in the system. These factors may be grouped in the following categories:

1. Economics — The approximate amount of money available for the development of the data system is usually the overriding constraint on system design. A thorough analysis of the problem, such as we have described, is primarily aimed at obtaining maximum performance from the proposed system within the available funds.

2. Physical Environment — The geographical and climatological regime in which the system must operate is a factor that automatically rules out certain approaches. Included in this category are considerations of morbidity, or the ease with which the measuring system is available for relocation.

3. Personnel — The training, capabilities, and number of personnel required to

operate and install the system can have a significant effect on system design philosophy and procedures.

4. Interfaces of the Meteorological System with Other Functions — A meteorological data collection system designed for furnishing information to an air traffic controller entails significantly different problems from one designed for study of atmospheric pollution around a city.

With the information developed from analysis of these factors we can establish system functional objectives consistent with the broad application requirements, technological state-of-the-art, and so-called "practical" limitations on the system design and development. These functional objectives properly described with their relationship to one another, form the basis for more detailed determination of the specific requirements for each of the subsystems: sensors, communications, data processing, and display.

The final phase of the design study involves specification of requirements for each subsystem and its important components. For the sensor subsystem characteristics such as the following must be specified:

1. Form of the sensor outputs (digital or analog).
2. Accuracy, resolution, and range of the sensor readings.
3. Required ruggedness and reliability, including protection from the environment, and electrical and mechanical functioning.

For the communication subsystem, major considerations will be:

1. Function of the communication system over fixed-wire channels or by radio telemetry.
2. Necessary bandwidth as determined both by the data rates generated from the meteorological measuring instruments and by reliability considerations.
3. Coding techniques, particularly whether signals are to be transmitted in analog or digital form.
4. Power requirements.
5. Physical maintenance requirements.

For data processing subsystem we must specify:

1. Which parts of the processing are to be automatic and which are to be handled manually.
2. Appropriate processing speeds and memory requirements for the automatic data handling.
3. Necessary automatic (real-time) computer inputs, if any.
4. Appropriate forms of recording data not directly entered into the computer, such as punch cards, punch paper tapes, or strip-chart graphs.
5. Signal conversion operations, such as analog-to-digital conversion.
6. Computational requirements, such as objective forecast models, turbulent diffusion models, and atmospheric statistics.
7. Data recording requirements and appropriate forms, such as printed copy, magnetic tape, punch cards, etc.

For the information display subsystem of our data collection system, appropriate forms and numbers for display devices and materials must be indicated. These include:

1. Necessity for automatic moving displays, such as cathode ray tubes and various types of projection systems.
2. Printed record requirements.
3. Automatic alarm signal requirements.
4. Physical location and number of displays required.

HYPOTHETICAL METEOROLOGICAL INFORMATION SYSTEM

The system analysis techniques discussed above can be illustrated by a specific example. Let us consider a rather comprehensive system of measurements over a reference area of a few hundred square miles; say, a 20-mile by 20-mile square. In addition to knowing the general regional weather conditions that affect the area in which this square is located, we desire information on the local meteorology (mesoscale); further, we are interested in analysis of the turbulence structure on certain sub-regions of this square, say of about a few hundred yards on a side (microscale). To bring out all facets of the problem, we will assume stringent performance requirements for the data system. Such requirements are frequently met in missile launch control and chemical weapons testing, for example. More to the point, we believe that through careful preliminary planning such scope and flexibility can be achieved at no more cost than that of many current systems having restricted information-gathering and processing power.

We will suppose that we are interested in obtaining all of the meteorological data, including that pertaining to the turbulence structure, in digested form essentially instantaneously (in real-time). Hence we adopt the premise that our measurements will read directly into an on-line general-purpose digital computer, which will then furnish printed and moving displays. In addition we will assume that all data are to be recorded for future research and examination.

Macroscale weather data are to be entered manually into the computer in order to furnish information on general weather conditions to be expected. Mesoscale data acquired from our local meteorological instrumentation system are to be used for tracing and predicting the gross trajectory of various air contaminants. This tracing will be further refined through knowledge of the atmospheric turbulence structure obtained from our microscale measurements. We will refer to the measurements pertaining to mesoscale data as the meso-network, or simply mesonet, and those from the microscale measurements as the micronet.

The micronet must furnish information concerning temperature, relative humidity, wind speed, and wind direction at selected points over the reference area. These measurements must be made at sampling frequencies and over time intervals consistent with the form of the power spectrum characteristic of each parameter. In particular, the key to turbulence structure is in wind-direction fluctuations. Experience shows that frequencies up to 2 or 3 cycles per second in the wind-direction vector frequently contain significant energy. To measure this portion of the wind energy spectrum, wind direction must be sampled at a rate of, say, 10 times per second, at least during those intervals of time when we are seriously concerned with the effects of small-scale atmospheric turbulence. Temperature and relative humidity measurements will be made once per minute throughout our network.

Our mesonet will be designed to furnish spatial measurements of atmospheric pressure, temperature, temperature gradient, and net radiation, in addition to the wind measurements. An effective mesonet might consist of 20 to 25 sounding stations capable of providing meteorological data up to altitudes of 3,000 feet. Each sounding would furnish measurements of vertical distribution of ambient temperature, relative humidity, wind speed, and wind direction at 100-foot intervals. In addition a single net radiation sensor could be located at each sounding station. Although measurements of temperature and relative humidity would be obtained by direct measurement from the sounding instrument package, wind data must be computed from successive position information. We further envision that these position data are provided by simultaneous automatic tracking of perhaps five sounding packages. The location of the individual measurement stations is shown in Figure 4; a mirror image of this configuration, aligned along the opposite direction, is not shown in the figure. The micrometeorological measurement network is contained within the triangular array.

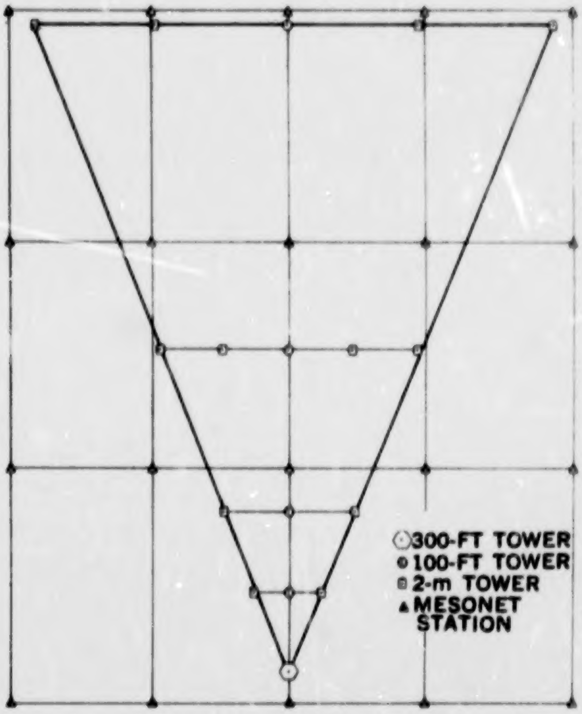


Figure 4 — Hypothetical Networks

The instrumentation subsystem associated with these networks must be capable of supplying reliable data to support two measurement objectives: (1) to obtain operational data required for routine estimates of pollution levels, and (2) to obtain the measurements that provide background information necessary for analysis and research. Often these requirements are significantly different from the operational requirements, both in the quantity of data required and in the timeliness of the information. The proper development of an automatic data acquisition involves the fusion of these differing needs into a well-organized system design.

BASIC MEASUREMENT REQUIREMENTS

The results of a synthesis of these requirements mentioned above for both the operational needs and for a research program are summarized in Table 2. Ten levels of measurement are provided for a 300-foot tower, eight levels of measurement for the 100-foot towers, and a single level for the 2-meter towers. The number of sensors of each type and the total number of sensors are given for each network. In addition to the number of sensors, the maximum sampling rate for each type of sensor is of vital concern to the systems engineer. The highest sampling rates are associated with bi-vane measurements of the horizontal and vertical components of the wind vector. Under normal operation, these instruments would be sampled at the rate of once per second (1/sec) with intermittent periods of the high data sampling. Aerovane measurements of wind speed and direction taken at the 2-meter towers are sampled once per second (1/sec). Temperature, temperature gradient, and dew point instruments can, however, be sampled as slowly as once per minute (1/60). This set of measurements generates our micronet data.

Table 2 — Measurement Configuration for Hypothetical System

MICRONET

Parameter	One 300-Foot Tower		Four 100-Foot Towers		Twelve 2-Meter Towers	
	No. of Sensors	Max Sampling Freq, Sec $^{-1}$	No. of Sensors	Max Sampling Freq, Sec $^{-1}$	No. of Sensors	Max Sampling Freq, Sec $^{-1}$
Temperature	10	1/60				
Dewpoint	10	1/60	28	1/60		
Temp. Gradient			24	1/60		
Radiation	1	1/60				
Azimuth Wind	4	10	8	10		
Elevation Wind	4	10	8	10		
Wind Speed	10	10	8	10	12	1
Wind Direc.	6	10			12	1
Total	45		76		24	

MESONET

Parameter	No. of Sensors	Max Sampling Freq, Sec $^{-1}$
Temperature	15	1/12
Dewpoint	15	1/12
Azimuth Angle	15	1/12
Elevation Angle	15	1/12
Time	15	1/12
Total	75	

SUMMARY

	No. of Sensors	Max Sampling Freq, Sec $^{-1}$
	48	10
	24	1
	75	1/12
	73	1/60
Total	220	

To produce a reasonably complete description of the mesoscale features of the atmosphere, rocket- or balloon-launched instrument packages are timed to transmit temperature and dew point data at the rate of five times per minute. Azimuth and elevation angles of the package are obtained by ground tracking instruments. Table 2 also shows the measurement program for the mesonet.

A total of 220 separate meteorological measurements are provided by these networks. Equally important to the system analysis is the summary of various categories of sampling rates. Only about one third of the sensors are sampled at rates of once per second or greater, with the higher rates occurring only intermittently. This information must be evaluated along with prescribed system timing requirements to develop initial estimates of the system data loads.

TIMING REQUIREMENTS AND DATA RATES

The combination of operational and research requirements imposes rigid timing constraints on performance of the measurement program necessary for pollution prediction. The next stage in the development of the system is to investigate the effect of these constraints on the sequences of measurements and the necessary time relationships between major system functions. These considerations are at the heart of the real-time operational problems associated with on-line computer systems such as we are considering. We again consider our hypothetical networks as the basis for illustrating the development of system specifications.

It is convenient to consider a typical 8-hour period, which we divide into two principal operational phases. The first phase consists of preparations and includes such activities as subsystem activation, system checkout, and preliminary atmospheric sampling. When a serious pollution problem is foreseen, we move to a second-phase measurement program, gathering more data, particularly on small-scale turbulence structures.

The timing requirements for such a program are illustrated in Figure 5. During phase one, 10-minute periods of high-frequency sampling are scheduled at hourly intervals along with mesonet releases 15 minutes in length. At these peak periods, data rates will be approximately 6600 bits/sec where the estimates of data rates are based on a binary code with 13 bits/word. This 13-bit word format assumes only data information and does not include bit requirements for control purposes. The addition of identification and control data to the message format can be included in a final evaluation of data rates by proportional scaling of the values shown in the graph. These data rates also will require modification as a consequence of the method of sensor interrogation employed and the number of re-transmissions used for checking and validation.

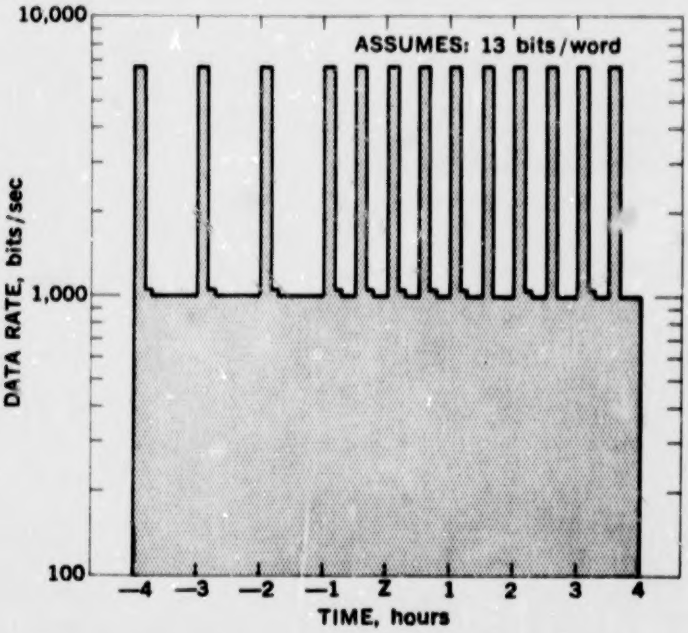


Figure 5 — Meteorological Information Data Rates

During the second phase, the high-data-rate periods are repeated at half-hour intervals. The frequency release of mesonet sounding package, however, decreases toward the end of this phase. During low-data-rate sampling, the curves show a steady system data load of approximately 1000 bits/sec.

Another aspect of the data load picture, particularly from the standpoint of data recording and storage requirements, is illustrated in Figure 6. This curve shows the time profile of accumulated meteorological information. If all data from our hypothetical system are recorded during the single 8-hour experiment, a total of 5.6 million data words will be accumulated.

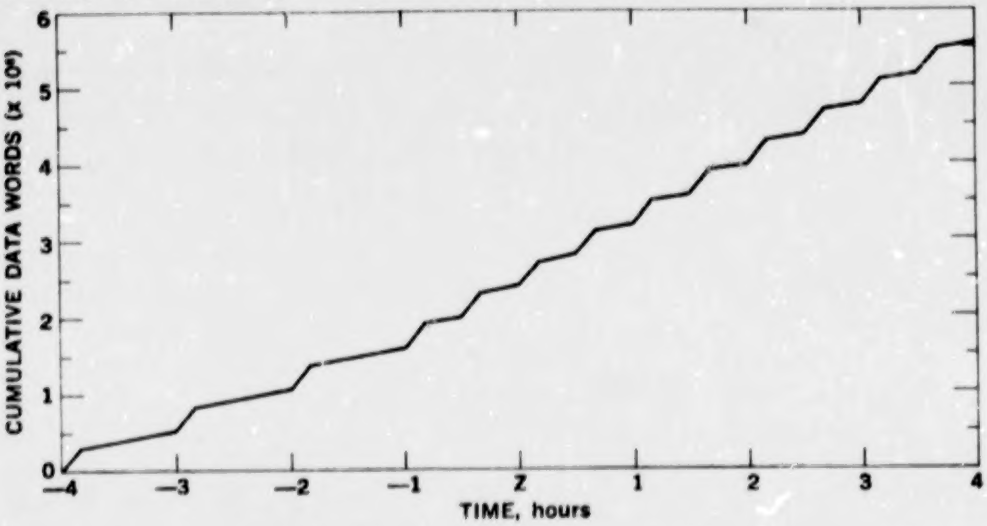


Figure 6 — Cumulative Data Words Versus Time

This type of information and the procedure followed in its development are essential steps in the preliminary phases of the system design. The estimates thus derived may be modified at a later date but serve the very important role of providing a basis upon which initial specifications of the sensor, telemetry, data processing, and display subsystems may be made.

Let us review here the functional requirements of this system. We are interested in obtaining all of the meteorological data, including that pertaining to the turbulence structure, in digested form essentially instantaneously (in real-time) to facilitate short-term pollution predictions. Our measurements will read directly into an on-line general-purpose digital computer. The computer will then furnish printed and moving displays. In addition we have assumed that all data are to be recorded for future research and examination.

SUBSYSTEM CHARACTERISTICS

Analysis of these functional requirements and data rates subject to the above design considerations leads to the following description of basic characteristics for the four major subsystems.

1. Sensors — Table 3 presents a summary of the number and required accuracy for each type of sensor to be installed in the hypothetical system. This information may serve as a basis for further investigation of appropriate specifications and for surveying commercially available sensors. Results of this survey may indicate that accuracy requirements must be relaxed if commercially available sensors are to be used.

Table 3 — Summary of Sensor Subsystem Specifications

Parameter	Type	Number	Accuracy	Minimum Digitizing Interval
Temperature (Temperature Gradient)	Thermocouples	34	± .02°C	1.0 Min
Wind Speed	Anemometer	30	5% of Wind Speed or 2 ft/sec	0.1 Sec
Horizontal Wind Direction	Wind Vane	18	1°	1.0 Sec
Horizontal Vertical } Wind Direction	Bi-Vane	24	1°	0.1 Sec
Dewpoint	Dewcell	38	1%	1.0 Min
Net Radiation	Radiometer	1	2%	
	Total	145		

2. Telemetry — To transmit the measurements from remote sensor locations to the central location for processing, a telemetry system is required. Accuracy and reliability dictate the need for digitizing of all data to be transmitted over any significant distance. The data rates indicate a peak transmission rate of approximately 6600 bits/sec. Control and checking requirements may increase the necessary telemetering capacity by a factor of 2 to 4. These factors strongly indicate the use of r-f data links. If we allow an additional factor of 2 to 4 for adequate modulation and signal-to-noise ratio, a bandwidth of from 50 to 100 kc will be required for a single r-f communication channel.

Radio transmission problems peculiar to the area over which the system is to operate may dictate a requirement for the use of relay stations. In the hypothetical system it is assumed that one relay station is required because of obstructions in the center of the area. Separate channels are required for transmission and reception and for communication between the relay station and the central data collection and between the relay and the remote sensor location. This leads to the requirement for five separate 50 to 100 kc channels. These requirements are listed in Table 4.

Table 4 — Summary of Communications Subsystem Specifications

Type of Transmission	r-f
Frequency	200-250 Megacycles
Number of Channels	5
Bandwidth	50-100 kc
Number of Relay Stations	1

3. Data Processing — The central control element of the meteorological measurement system is a digital computer. The need for this element to be a general-purpose computer is of major importance, since flexibility is a primary objective of the hypothetical meteorological measurement system. This flexibility cannot be achieved with a special-purpose computer with wired-in programs. The computer controls the reading sequence of the meteorological sensors, actuates the digitization and communication operations, computes necessary control parameters and test decision criteria, and operates printed and moving display outputs. In addition the computer edits and records all raw measurements on magnetic tape. With the speeds and asynchronous control available on small computers today it is reasonable to suppose that diffusion predictions

and research calculations are time-shared concurrently with these operations. Table 5 presents a listing of basic features of the computer.

Table 5 — Summary of Data Processing Subsystem Specifications

DIGITAL COMPUTER CHARACTERISTICS

- 8,000 Words of Memory (Expandable to 16,000)
- 20 Microseconds Memory Cycle Time
- Magnetic Tape Units For Recording
- Auxiliary Drum Memory
- Automatic Interrupt
- Real-Time Clock
- Simultaneous Computing With Input and Output
- Line Printer or Flexowriter
- Paper Tape — Reader-Punch

4. Display — The display requirements are derived from the need for real-time visual presentation of the wind profile over the entire area. To accomplish this we propose a cathode ray tube display (CRT), which displays the wind profile as vectors the length of which are measures of the wind speed and the direction of which indicate wind directions. Such a vector will appear for each tower location at a fixed sensor height and the height level chosen will be under control of the observer.

In addition to the wind profile it is desirable to observe the trends of the meteorological parameters. From these trends it would be possible to make short-term predictions of these parameters. Commercially available trend recorders fulfill this requirement.

In many cases there are some sensors which are of vital importance to the observer. These important sensors require so-called "go-no-go" displays which display a red light if the measurement being reported is outside tolerable limits, green if it is within tolerance, and a third color if something appears wrong with the sensor.

Additional printed display of parameters and summaries of computed quantities can be furnished through an electric typewriter. These requirements are listed in Table 6.

Table 6 — Summary of Display Subsystem Specifications

Type	Number
Cathode Ray Tube Display	1
Trend Recorders	50
Automatic Go-No-Go Displays	60
Automatic Alarm	

ACKNOWLEDGMENTS

Three staff members of Systems Research Laboratory of Geophysics Corporation of America made important contributions to the contents of this paper. Mr. David D. Dix is responsible for much of the material on basic system design. Mr. David Farrell and Mr. Paul Morgenstern developed the data rates and other details of the hypothetical system.

REFERENCES

Brown, R. M., 1959: An Automatic Meteorological Data Collecting System, *J. Geophys. Res.*, *64*, 2369-2372.

Cramer, H. E., F. A. Record, J. E. Tillman, and H. C. Vaughan, 1961: Studies of the Spectra of the Vertical Fluxes of Momentum, Heat, and Moisture in the Atmospheric Boundary Layer, *Annual Report* (Contract DA-36-039-SC-80209), Mass. Inst. of Tech., 130 pp.

Haugen, D. A., R. F. Myers, and J. H. Taylor, 1962: Design and Development of a Micrometeorological Data Observing and Processing System for Air Pollution Applications at Cape Canaveral and Vandenberg Air Force Base, Paper Presented at Fourth Conference on Applied Meteorology, American Meteorological Society, Hampton, Virginia, 10-14 September 1962, 22 pp.

Moses, H. and F. C. Kulhanek, 1962: Argonne Automatic Meteorological Data Processing System, *J. Appl. Meteor.*, *1*, 69-80.

Myers, R. F., 1956: A Weather Information Telemeter System, *Bull Amer. Meteor. Soc.* *37*, 108-117.

Islitzer: Personal Communication.

Dr. Oscar J. Balchum
Hastings Associate Professor of Medicine
University of Southern California School of Medicine, Los Angeles
and
Dr. Frank J. Massey
Associate Professor of Biostatistics
University of California at Los Angeles

SUMMARY

If we are to determine the effects of air pollutants on the human respiratory system, we must know more about the physical and mechanical properties of the chest and lungs. Techniques for measuring air pollution effects must be sensitive, accurate, and repeatable. Application of computer analysis to these measurements should then yield useful and reliable information on data acquisition systems for physiological studies. Investigations now under way are described: the parameters measured, the recording and coding of data, and the methods of analysis.

DATA ACQUISITION SYSTEMS IN PHYSIOLOGY[*]

INTRODUCTION

During the past decade researchers have suspected that chronic exposure to low levels of foreign gases and aerosols is a factor in the etiology of chronic respiratory disease. Investigations of the physiological reactions of the lungs of animals and man to low concentrations of particulates, aerosols, and gases, singly and in combination, have begun only recently. Although measurement systems are being developed for these studies, the techniques developed thus far are not sufficiently sensitive, accurate, or repeatable. Because our knowledge of the properties of the chest wall and lungs is meager, it has been difficult to measure small degrees of response to low levels of irritants. Both increased knowledge and improved instrumentation are required for effective investigations of the physiological effects of pollutants in the ambient air.

In work now under way, various properties of the respiratory system are measured and recorded in forms suitable for computer processing. This presentation describes the parameters measured, the recording and coding of data, and the methods of analysis.

MEASUREMENTS

At present the physiological reactions of the lungs are described by data obtained in a battery of tests.[1] These tests may be classified on the basis of the property of the respiratory system being measured.

VOLUMES OF THE LUNGS (Figure 1)

Vital Capacity (VC)—the greatest amount of air that can be exhaled after the deepest possible inhalation. Vital capacity is measured by means of a 13½-liter Collins

---

[*] Supported in part by the grant for the National Institutes of Health (AP207) and by a contract with the Air Pollution Division of the United States Public Health Service (PH 86-62-162).

Spirometer. Values are related to sex, age, and height; prediction nomograms have been established from these values.

Examples: Men — 39-50 years, VC = 3450 ml
Standard Deviation (s) = 980 ml
Coefficient of Variation (CV)* = 28%
Standard Error (SE) = 80 ml

Women — 40-67 years, VC = 2880 ml
s = 630 ml
CV = 22%
SE = 60 ml

*Coefficient of variation, or the SD expressed as a percentage of the mean.

Functional Residual Capacity (Resting Level or FRC) — the amount of air in the lungs at the end of an ordinary exhalation.

Examples: Men — 2180 ml
s = 690 ml
CV = 32%

Women — 1830 ml
s = 420 ml
CV = 23%

Residual Volume (RV) — the amount of air still remaining in lungs after the deepest possible exhalation, measured by the helium dilution method. The apparatus consists of a closed circuit, with a spirometer, a CO_2 absorber, and a helium thermoconductivity cell in series. The circuit is filled with 15 percent helium in air. The patient wears a nose clip and is attached to the two-way valve of the apparatus by a rubber mouthpiece. At the end of an ordinary exhalation, he is connected to the circuit and then breathes the 15 percent helium in air mixture.

Oxygen is supplied to the circuit at the same rate at which it is consumed. After helium has diffused into the lungs so that equilibrium or a plateau of the concentration curve has been achieved, the final concentration of helium is read on the galvanometer.

Calculation of FRC: (Initial He Conc) (Vol of Circuit) = (Final He Conc) (Vol of Circuit + FRC).

The only unknown, the FRC, can then be calculated. Residual Volume (RV) is obtained by subtracting the Expiratory Capacity (obtained from the spirogram) from the FRC.

Examples: Men — 1140 ml
s = 430 ml
CV = 38%

Women — 995 ml
s = 280 ml
CV = 28%

Total Capacity — = Vital Capacity + Residual Volume.

Examples: Men — 4590 ml
s = 1300 ml
CV = 28%

Women —3880 ml
s = 800 ml
CV = 21%

$$RV/TC\ \% = \frac{\text{Residual Volume}}{\text{Total Capacity}} \times 100$$

Examples: Men —24.9%
s = 4.7%
CV = 19%

Women —25.7%
s = 4.7%
CV = 18%

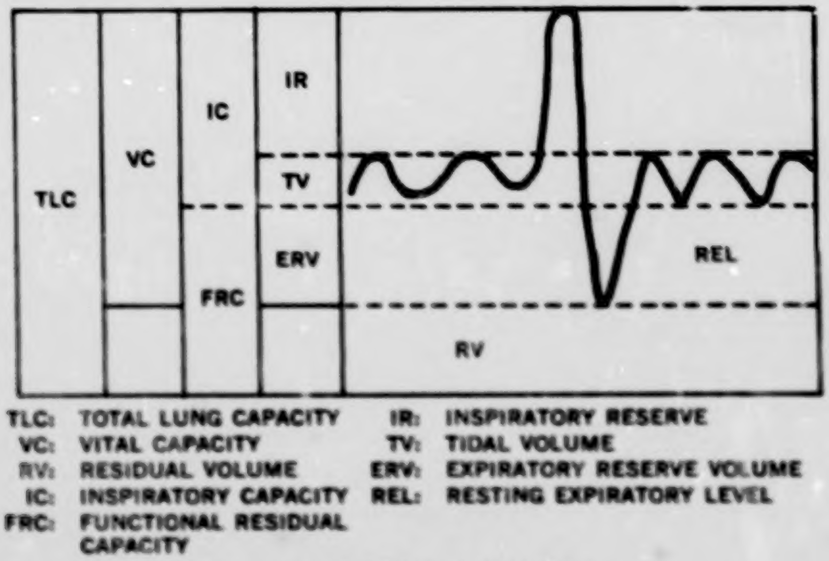

IR
IC
VC
TV
TLC
ERV
REL
FRC
RV
TLC: TOTAL LUNG CAPACITY
VC: VITAL CAPACITY
RV: RESIDUAL VOLUME
IC: INSPIRATORY CAPACITY
FRC: FUNCTIONAL RESIDUAL CAPACITY
IR: INSPIRATORY RESERVE
TV: TIDAL VOLUME
ERV: EXPIRATORY RESERVE VOLUME
REL: RESTING EXPIRATORY LEVEL

Figure 1 — Definition of Lung Volumes.

VENTILATION

Maximum Breathing Capacity (MBC) — the greatest amount of air that can be inhaled and exhaled into and out of the spirometer. The subject is asked to breath in and out as rapidly and as deeply as necessary in order to ventilate as much air as possible. This is done at a rate above 80 breaths per minute for 12 seconds. The resulting volume is multiplied by 5 to obtain MBC in liters per minute.

Examples: Men —103 l/min
s = 32 l/min
CV = 28%

Women —89 l/min
s = 20 l/min
CV = 22%

Forced Expiratory Volume (FEV or Timed Vital Capacity) — the amount of air exhaled (upon command) as rapidly and completely as possible after the deepest possible breath, with the drum of a spirometer rotating at 600 or 960 mm/min. The FEV can be measured from the spirogram.

FEV_1: the liters of air exhaled during the first second of a forced expiratory volume.

FEV_3: the liters of air exhaled during the first three seconds of the FEV.

Volumes are corrected to BTPS, or body temperature (37°C), saturated. If a lung volume was measured at a spirometer temperature of 25°C, and 750 mm Hg atmospheric pressure, the correction would be:

$$\text{Volume}_{BTPS} = \text{Vol} \; \frac{(273 + 37)}{(273 + 25)} \; \frac{(750 - 24)}{(750 - 47)}$$

24 and 47 mm Hg are the water vapor tensions at room and body temperatures, respectively.

The FEV and FEV_3 are also used as a percentage of the total FEV:

$$FEV_1, \% = \frac{FEV_1}{FEV} \times 100$$

Average FEV_1 = 83% (Range 70-90%)

Average FEV_3 = 97% (Range 90-100%)

Example: $FEV_1 = 3.93$ l, $s = 0.67$ l
$FEV_1 = 82.0\%$, SE = 1.12
$FEV_3 = 97.6\%$ (Range 92-100%)

Maximal Mid-Expiratory Flow Rate (MMF)

MMF $_{25-75}\%$ = the rate of flow of air during the middle 50 percent of the FEV, expressed in liters per second.

Normal values not well established.

Example: 4.49 l/sec.
s = 1.3 l/sec
SE = 0.25 l/sec

Peak Flow Rate — the maximal or peak rate of air flow measured during a rapid or blastlike exhalation (after a deep inhalation) into a Wright Peak Flow Meter. The meter consists of an encased light vane that rotates and stops at the point of peak flow. The dial of the instrument is calibrated in liters per minute.

Other instruments (Puffmeter, Pneumotachygraph) consist of small resistances, the pressure drop across which is proportional to rate of air flow. The Puffmeter has as its resistance a porous cuplike grinding wheel, and the Pneumotachygraph a 400-mesh stainless steel screen. Recordings are made by use of amplifier and strip-chart recorder units of suitable frequency response.

LUNG MIXING

Rate of fall of helium concentration during performance of residual volume measurement.

7-Minute Lung Nitrogen Washout — Oxygen is inhaled by the subject, who breathes in a normal fashion. The nitrogen concentration is recorded during each exhalation and falls as the subject continues to breathe oxygen. Normal air distribution and lung mixing will result in a nitrogen concentration less than 1.5 percent in 7 minutes, with no appreciable increase during the performance of a maximal exhalation at this time.

LUNG DIFFUSION

The rate of passage of a tracer gas (0.05% carbon monoxide) from the air sacs (alveolae) of the lungs into the blood, as measured by an IR spectrophotometer. The gas is rapidly taken up but not released by the hemoglobin of the red cells and therefore exerts little back-pressure.

The concentration of CO breathed in and that exhaled, and the volume of gas mixture breathed are recorded during a period of 4 to 5 minutes. During the latter half of such a period the rate of passage of CO from the inhaled CO-air mixture becomes constant, the CO concentration of the exhaled gas mixture forming a plateau on the record.

Indices Calculated

$$\text{Uptake of CO, \%} = \frac{(\text{Min. Vol})\ (\text{CO}_{\text{insp}} - \text{CO}_{\text{exp}})}{(\text{Min. Vol})\ (\text{CO}_{\text{insp}})}$$

Example: 51.1%, s = 4.66%

Diffusion Capacity, ml/min per mm Hg =

$$\frac{(\text{Min. Vol})\ (\text{CO}_{\text{insp}} - \text{CO}_{\text{exp}})}{(\text{End-Tidal Conc. CO})\ (\text{Barometric Pressure} - 47)}$$

Example: 23.3 ml/min per mm Hg
s = 4.93 ml/min per mm Hg

Mean difference between first and second paired estimates = 0.64, s = 3.12

AIRWAY RESISTANCE AND THORACIC GAS VOLUME BY BODY PLETHYSMOGRAPHY

Airway Resistance — the subject sits in an airtight box or plethysmograph. Box pressure (Box P) is recorded by means of a strain gage and rate of airflow by means of a pneumotachygraph screen while the subject pants at a rate of about 120 times per minute at the FRC level. A vector loop of flow rate (Y axis) is plotted against box pressure (X axis), and the slope of the long axis of the vector loop measured on a CRO.

A few seconds later, at the instant exhalation reaches the FRC level, a solenoid completely obstructs the tube between the patient's mouth and the pneumotachygraph screen while he is still panting.

Mouth pressure (Y axis) is plotted against box pressure (X axis) on the CRO. During the few seconds of complete obstruction, the pressure in the alveolae (air sacs) of the lung is considered to be in equilibrium with that in the mouth, and alveolar (ALVP) or lung pressure = mouth pressure.

$$\text{Resistance of Airways, cm } H_2O \text{ per l/sec} = \frac{\text{Alv P/Box P}}{\text{Flow Rate/Box P}}$$

Example: 1.5 cm H_2O per l/sec
s = 0.37 cm H_2O per l/sec

$$\text{Thoracic Gas Volume (TGV), liters} = \frac{970}{\Delta P \Delta V}$$

[Note: 970 cm H_2O is the atmospheric pressure minus water vapor tension at body temperature (37°C), the conditions in the lungs.]

Example: 2.97 l
s = 0.22 l
SE = 0.07

THE RECORDING AND CODING OF DATA

After computation with pencil, paper, and desk calculator the results of these pulmonary function tests are entered on coded forms suitable for punching on IBM cards. The analyses of these data have been programmed for the IBM 7094 computer of the Western Data Processing Center, Los Angeles. Figures 2 through 8 present example forms for coding the results of vital capacity and timed vital capacity, lung nitrogen washout, carbon monoxide lung diffusion, thoracic gas volume and airway resistance by body plethysmography, and residual volume. Forms are given also for recording of air pollutant concentration levels (Figure 7) and of objective signs detected upon examination of the chest and of symptoms obtained by the questioning of the patient (Figure 8). These forms are a part of the system now being used for recording data in a study of patients with chronic respiratory disease. Records are taken while the patients reside for days in a room supplied with ambient Los Angeles air, and again while this room is supplied with air filtered through absolute and activated charcoal filters, at the Los Angeles County General Hospital (USC).

Similar methods are being used in a second study for coding and entering data for transfer to IBM punch cards according to a planned format for later transfer to magnetic tape. Here data on occupation, smoking, exposure to lung irritants, respiratory symptoms, etc., and the results of physical and x-ray examination of the chest and of lung function tests are being recorded annually, in an effort to depict the course or natural history of individuals who are "normal" or bronchitic, or who are already emphysematous. A year-to-year comparison of any data can be programmed from this longitudinal clinical and physiological investigation into the development of chronic respiratory disease in man.

ANALYSIS AND INTERPRETATION

The data can be checked by a complete printout (Figure 9), with inspection for values that appear to deviate more widely than expected, and for missing values. A complete listing of variables is printed out (Figure 10), with tabulation of the variable numbers, variable names, the number of non-zero cases, the means, standard deviation, and high, low, and range of values. These two tabulations aid greatly in the detection of punching errors even though the card punching already has been verified, and in the identification of missing values or a wrong order of cards including those cards directing the programmed sequence of steps in the analysis. These tabulations also have been of great help in locating technical or measurement errors.

A histogram and cumulative frequency polygon (Figure 11) are used to describe the distribution of values and to give the percent of cases within the range of limits of the variable selected. It should be noted that any restrictions can be placed on the variables selected. These restrictions may include characteristics such as age, sex, or a limit of pulmonary function test result, etc.

A two-way plot or scatter diagram (Figure 12) with computation of the mean and standard deviation of each variable, the coefficient of correlation, and the equation of the line of regression is very informative. A two-way table (Figure 13) with or without restrictions enables obtaining various frequencies according to the ranks selected, and the estimates of variation (SD, Chi square).

Tables can be obtained with various restrictions on variables, with printout of the means and ranges of the values for each, and printout of individual values (Figure 14).

Row and column restrictions of any combination desired can be handled (Figure 15). Nested distribution tables (tables within tables) are useful in analyzing studies involving multiple variables (Figure 16). Analyses of data according to selected restrictions with computation of correlation coefficients and regression coefficients are available (Figure 17).

On-line transmission of instrument output signals directly to a computer or to a magnetic tape recorder is not yet in use. Exploration of the method is beginning, and the method has been used in the recording and analysis of vectorcardiograms. A "computer-spirometer" is available, having a readout of calculated values derived from the FEV, such as the VC and FEV_1, %. All these systems would be useful in surveying by spirometric methods large numbers of individuals for chronic respiratory disease.[4]

Pulmonary function testing is based upon the careful handling and instructing of patients in the breathing maneuvers desired, adherence to the conditions of each test, and a careful setting up of the instrument system; hence it demands trained personnel. On-line instrument-computer systems probably would be too costly for the usual hospital pulmonary function laboratory. Smaller systems for spirometry are probably feasible, since this is a commonly performed yet extremely informative measurement.

REFERENCES

1. Balchum, O. J. Instrumentation and Methods for Measuring the Physiological Effects of Air Pollution. ISA Biomedical Sciences Instrumentation Symposium, June 14-18, 1963, Los Angeles, California. Symposium Proceedings, Plenum Press, 227 West 17th Street, New York 17, New York.

2. Swann, H. E., Brunol, D., and Balchum, O. J. An Improved Method for Measuring Pulmonary Resistance in Guinea Pigs — To Be Published.

3. Brunol, D., Balchum, O. J., and Swann, H. E. Mechanics of the Chest and Lungs: Physical Basis for Pulmonary Resistance Measurement — To Be Published.

4. Balchum, O. J., Felton, J. S., Jamison, J. N., Gaines, R. S., Clarke, D. R., and Owan, T. A Survey for Chronic Respiratory Disease in an Industrial City. Amer. Rev. of Resp. Dis. 86:675, 1962.

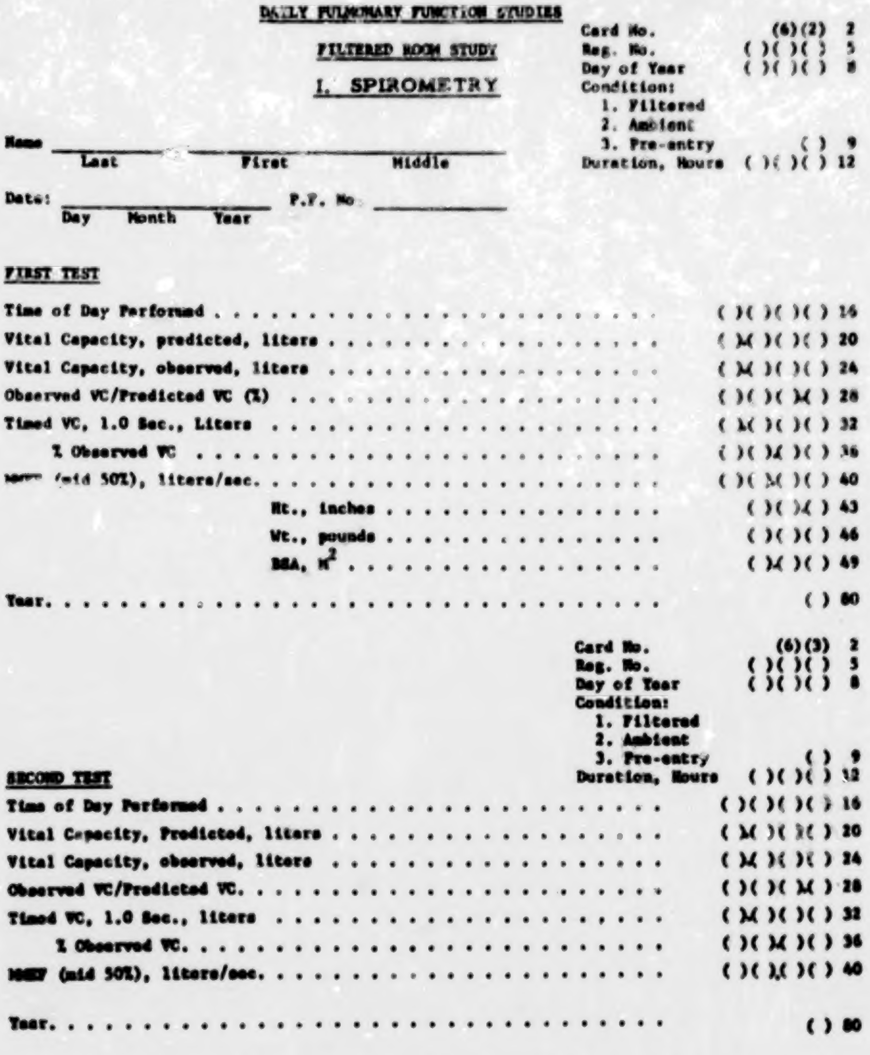
DISCUSSION

Asked whether any known correlation had been shown as a result of his studies, Dr. Balchum replied that insufficient data have been collected for practical statistical evaluation and that at this time his results are inconclusive. This possibly might be due to the low-level pollutant concentrations in the ambient air now used for evaluation. Dr. Balchum indicated that better correlation between pollutants and physiological effects might be expected when higher pollutant levels found during the California smog season are used for exposures.

When asked how subjects are obtained for the physiological research studies, Dr. Balchum indicated that the University hospital maintains a roster of approximately 300 respiratory patients who volunteer as subjects; also, new patients are solicited to serve as study subjects.

Dr. Zavon asked whether any physiological reactions other than respiratory response are being measured. The reply indicated that no other measurements, such as blood or urine analysis, are being attempted. Dr. Zavon pointed out that the chemical reaction of such air pollutant substances as 3-4 benzpyrene are being investigated, both in this country and abroad, for response in other portions of the human system.

A participant asked whether the ambient air used as test atmosphere for the patients is altered when it is passed through the spyrometer during tests. Dr. Balchum replied that although no tests have been made to determine whether the NO_x or oxidant levels are reduced in passing through the spyrometer, he believes that because of the relatively short time that the patient breaths through this transducer such losses are not a significant factor for a 24-hour test period.



DAILY PULMONARY FUNCTION STUDIES

FILTERED ROOM STUDY

I. SPIROMETRY

Card No.	(6)(2)	2
Reg. No.	( )( )( )	5
Day of Year	( )( )( )	8
Condition: 1. Filtered 2. Ambient 3. Pre-entry	( )	9
Duration, Hours	( )( )( )	12

Name ______ ______ ______
Last First Middle

Date: ______ P.F. No ______
Day Month Year

FIRST TEST

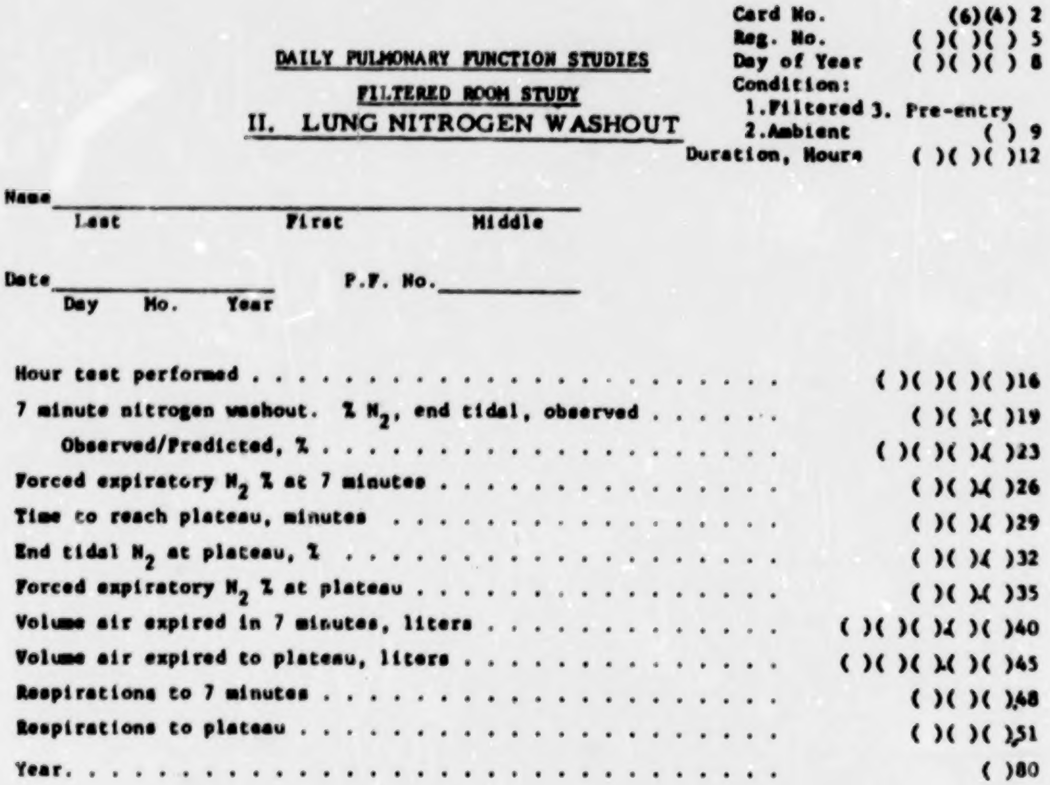
Time of Day Performed	( )( )( )( )	16
Vital Capacity, predicted, liters	( ).( )( )( )	20
Vital Capacity, observed, liters	( ).( )( )( )	24
Observed VC/Predicted VC (%)	( )( )( ).( )	28
Timed VC, 1.0 Sec., Liters	( ).( )( )( )	32
% Observed VC	( )( ).( )( )	36
MMEF (mid 50%), liters/sec.	( )( ).( )( )	40
Ht., inches	( )( ).( )	43
Wt., pounds	( )( )( )	46
BSA, M^2	( ).( )( )	49
Year	( )	80

Card No.	(6)(3)	2
Reg. No.	( )( )( )	5
Day of Year	( )( )( )	8
Condition: 1. Filtered 2. Ambient 3. Pre-entry	( )	9
Duration, Hours	( )( )( )	12

SECOND TEST

Time of Day Performed	( )( )( )( )	16
Vital Capacity, Predicted, liters	( ).( )( )( )	20
Vital Capacity, observed, liters	( ).( )( )( )	24
Observed VC/Predicted VC.	( )( )( ).( )	28
Timed VC, 1.0 Sec., liters	( ).( )( )( )	32
% Observed VC.	( )( ).( )( )	36
MMEF (mid 50%), liters/sec.	( )( ).( )( )	40
Year	( )	80

Figure 2 — Data Form: Spirometry Tests



Card No. (6)(4) 2
Reg. No. ( )( )( ) 5
Day of Year ( )( )( ) 8
Condition:
1.Filtered 3. Pre-entry
2.Ambient ( ) 9
Duration, Hours ( )( )( )12

DAILY PULMONARY FUNCTION STUDIES

FILTERED ROOM STUDY

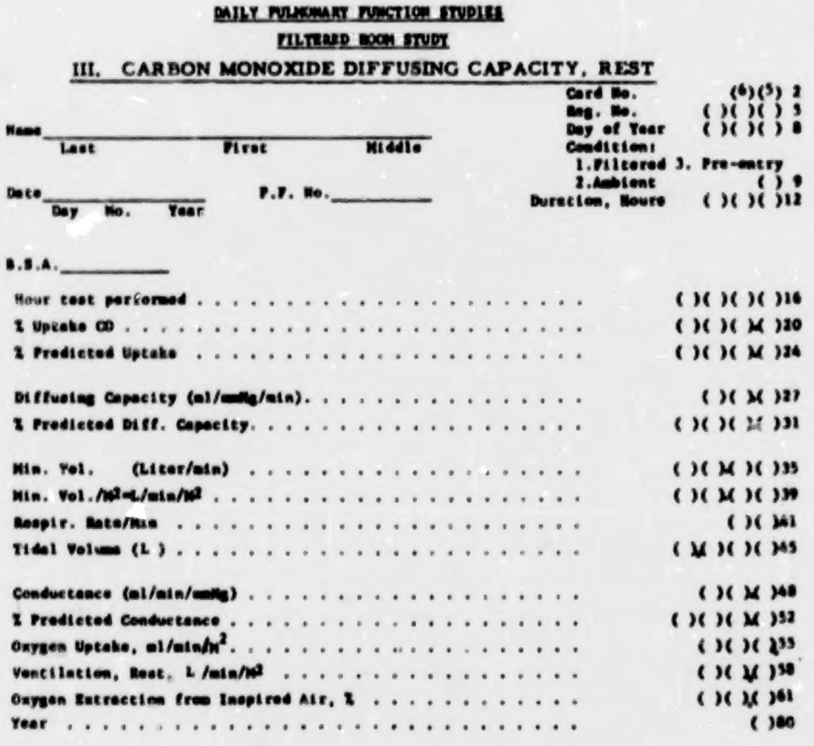
II. LUNG NITROGEN WASHOUT

Name____________________________
Last First Middle

Date______________ P.F. No.__________
Day Mo. Year

Hour test performed . . . . ( )( )( )( )16
7 minute nitrogen washout. % N_2, end tidal, observed . . . . ( )( ).( )19
Observed/Predicted, % . . . . ( )( )( ).( )23
Forced expiratory N_2 % at 7 minutes . . . . ( )( ).( )26
Time to reach plateau, minutes . . . . ( )( ).( )29
End tidal N_2 at plateau, % . . . . ( )( ).( )32
Forced expiratory N_2 % at plateau . . . . ( )( ).( )35
Volume air expired in 7 minutes, liters . . . . ( )( )( ).( )( )40
Volume air expired to plateau, liters . . . . ( )( )( ).( )( )45
Respirations to 7 minutes . . . . ( )( )( ).48
Respirations to plateau . . . . ( )( )( ).51
Year. . . . . ( )80

Figure 3 — Data Form: Lung Nitrogen Washout



DAILY PULMONARY FUNCTION STUDIES

FILTERED ROOM STUDY

III. CARBON MONOXIDE DIFFUSING CAPACITY, REST

Card No. (6)(5) 2
Reg. No. ( )( )( ) 5
Day of Year ( )( )( ) 8
Condition:
1.Filtered 3. Pre-entry
2.Ambient ( ) 9
Duration, Hours ( )( )( )12

Name____________________________
Last First Middle

Date______________ P.F. No.__________
Day Mo. Year

B.S.A.__________

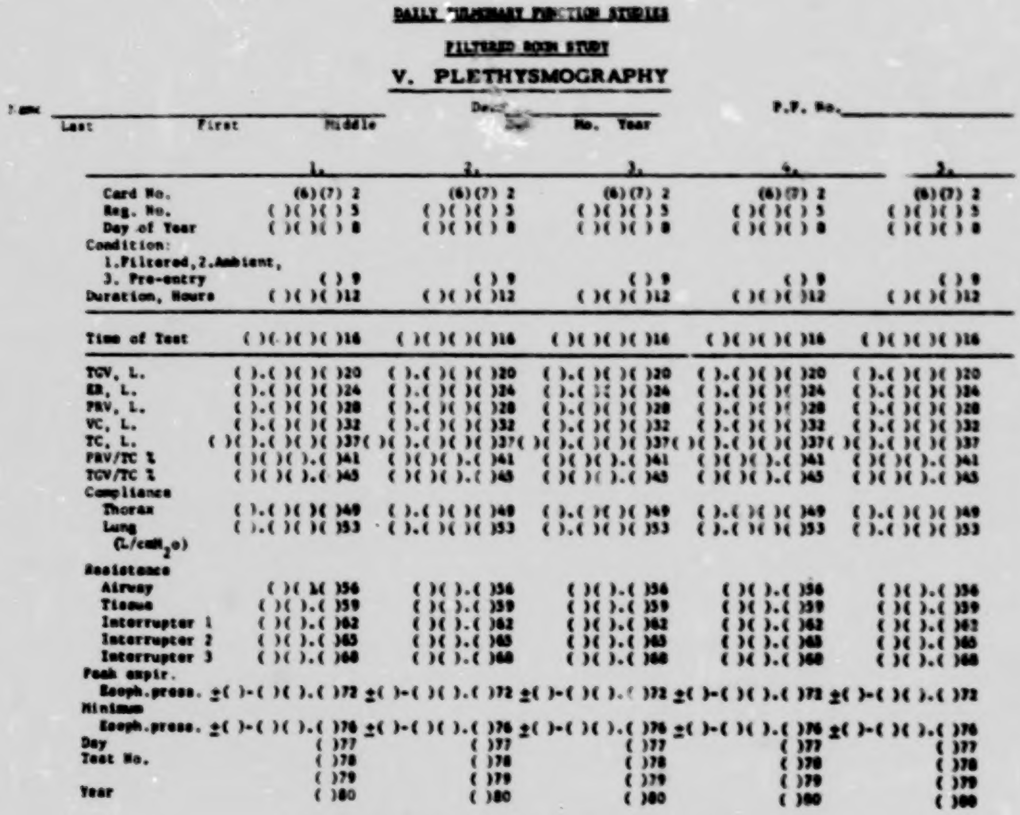
Hour test performed . . . . ( )( )( )( )16
% Uptake CO . . . . ( )( )( ).( )20
% Predicted Uptake . . . . ( )( )( ).( )24

Diffusing Capacity (ml/mmHg/min). . . . . ( )( ).( )27
% Predicted Diff. Capacity. . . . . ( )( )( ).( )31

Min. Vol. (Liter/min) . . . . ( )( ).( )( )35
Min. Vol./M²=L/min/M² . . . . ( )( ).( )( )39
Respir. Rate/Min . . . . ( )( )41
Tidal Volume (L) . . . . ( ).( )( )( )45

Conductance (ml/min/mmHg) . . . . ( )( ).( )48
% Predicted Conductance . . . . ( )( )( ).( )52
Oxygen Uptake, ml/min/M^2. . . . . ( )( )( )55
Ventilation, Rest, L /min/M^2 . . . . ( )( ).( )58
Oxygen Extraction from Inspired Air, % . . . . ( )( ).( )61
Year . . . . ( )80

Figure 4 — Data Form: Carbon Monoxide Diffusing Capacity, Rest



DAILY PULMONARY FUNCTION STUDIES

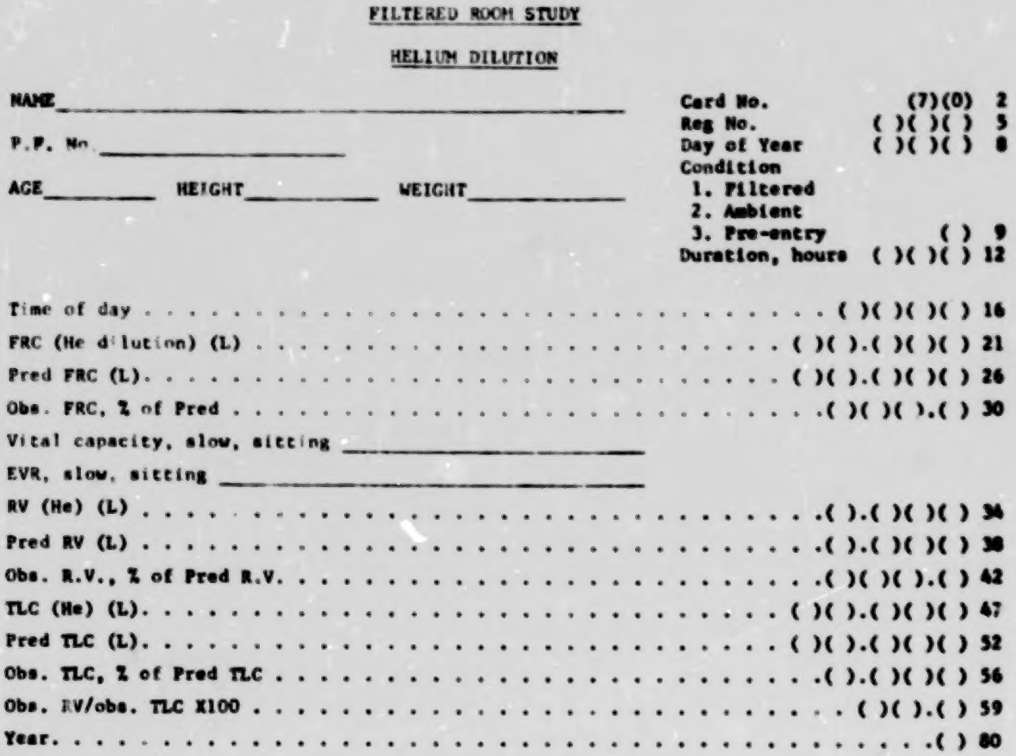
FILTERED ROOM STUDY

V. PLETHYSMOGRAPHY

Name ______________________ Date __________ P.F. No. __________
Last First Middle Day No. Year

	1.	2.	3.	4.	5.
Card No.	(6)(7) 2	(6)(7) 2	(6)(7) 2	(6)(7) 2	(6)(7) 2
Reg. No.	( )( )( ) 5	( )( )( ) 5	( )( )( ) 5	( )( )( ) 5	( )( )( ) 5
Day of Year	( )( )( ) 8	( )( )( ) 8	( )( )( ) 8	( )( )( ) 8	( )( )( ) 8
Condition: 1.Filtered,2.Ambient, 3. Pre-entry	( ) 9	( ) 9	( ) 9	( ) 9	( ) 9
Duration, Hours	( )( )( )12	( )( )( )12	( )( )( )12	( )( )( )12	( )( )( )12
Time of Test	( )( )( )( )16	( )( )( )( )16	( )( )( )( )16	( )( )( )( )16	( )( )( )( )16
TGV, L.	( ).( )( )( )20	( ).( )( )( )20	( ).( )( )( )20	( ).( )( )( )20	( ).( )( )( )20
ER, L.	( ).( )( )( )24	( ).( )( )( )24	( ).( )( )( )24	( ).( )( )( )24	( ).( )( )( )24
PRV, L.	( ).( )( )( )28	( ).( )( )( )28	( ).( )( )( )28	( ).( )( )( )28	( ).( )( )( )28
VC, L.	( ).( )( )( )32	( ).( )( )( )32	( ).( )( )( )32	( ).( )( )( )32	( ).( )( )( )32
TC, L.	( )( ).( )( )( )37	( )( ).( )( )( )37	( )( ).( )( )( )37	( )( ).( )( )( )37	( )( ).( )( )( )37
PRV/TC %	( )( )( ).( )41	( )( )( ).( )41	( )( )( ).( )41	( )( )( ).( )41	( )( )( ).( )41
TGV/TC %	( )( )( ).( )45	( )( )( ).( )45	( )( )( ).( )45	( )( )( ).( )45	( )( )( ).( )45
Compliance					
Thorax	( ).( )( )( )49	( ).( )( )( )49	( ).( )( )( )49	( ).( )( )( )49	( ).( )( )( )49
Lung (L/cmH$_2$o)	( ).( )( )( )53	( ).( )( )( )53	( ).( )( )( )53	( ).( )( )( )53	( ).( )( )( )53
Resistance					
Airway	( )( ).( )56	( )( ).( )56	( )( ).( )56	( )( ).( )56	( )( ).( )56
Tissue	( )( ).( )59	( )( ).( )59	( )( ).( )59	( )( ).( )59	( )( ).( )59
Interrupter 1	( )( ).( )62	( )( ).( )62	( )( ).( )62	( )( ).( )62	( )( ).( )62
Interrupter 2	( )( ).( )65	( )( ).( )65	( )( ).( )65	( )( ).( )65	( )( ).( )65
Interrupter 3	( )( ).( )68	( )( ).( )68	( )( ).( )68	( )( ).( )68	( )( ).( )68
Peak expir. Esoph.press.	±( )-( )( ).( )72	±( )-( )( ).( )72	±( )-( )( ).( )72	±( )-( )( ).( )72	±( )-( )( ).( )72
Minimum Esoph.press.	±( )-( )( ).( )76	±( )-( )( ).( )76	±( )-( )( ).( )76	±( )-( )( ).( )76	±( )-( )( ).( )76
Day	( )77	( )77	( )77	( )77	( )77
Test No.	( )78	( )78	( )78	( )78	( )78
	( )79	( )79	( )79	( )79	( )79
Year	( )80	( )80	( )80	( )80	( )80

Figure 5 — Data Form: Plethysmography



FILTERED ROOM STUDY

HELIUM DILUTION

NAME ______________________________

P.F. No. ____________________

AGE ________ HEIGHT __________ WEIGHT ____________

Card No.	(7)(0)	2
Reg No.	( )( )( )	5
Day of Year	( )( )( )	8
Condition 1. Filtered 2. Ambient 3. Pre-entry	( )	9
Duration, hours	( )( )( )	12

Time of day . . . . ( )( )( )( ) 16
FRC (He dilution) (L) . . . . ( )( ).( )( )( ) 21
Pred FRC (L). . . . . ( )( ).( )( )( ) 26
Obs. FRC, % of Pred . . . .( )( )( ).( ) 30
Vital capacity, slow, sitting ____________________
EVR, slow, sitting ____________________________
RV (He) (L) . . . .( ).( )( )( ) 34
Pred RV (L) . . . .( ).( )( )( ) 38
Obs. R.V., % of Pred R.V. . . . .( )( )( ).( ) 42
TLC (He) (L). . . . . ( )( ).( )( )( ) 47
Pred TLC (L). . . . . ( )( ).( )( )( ) 52
Obs. TLC, % of Pred TLC . . . .( ).( )( )( ) 56
Obs. RV/obs. TLC X100 . . . . ( )( ).( ) 59
Year. . . . .( ) 80

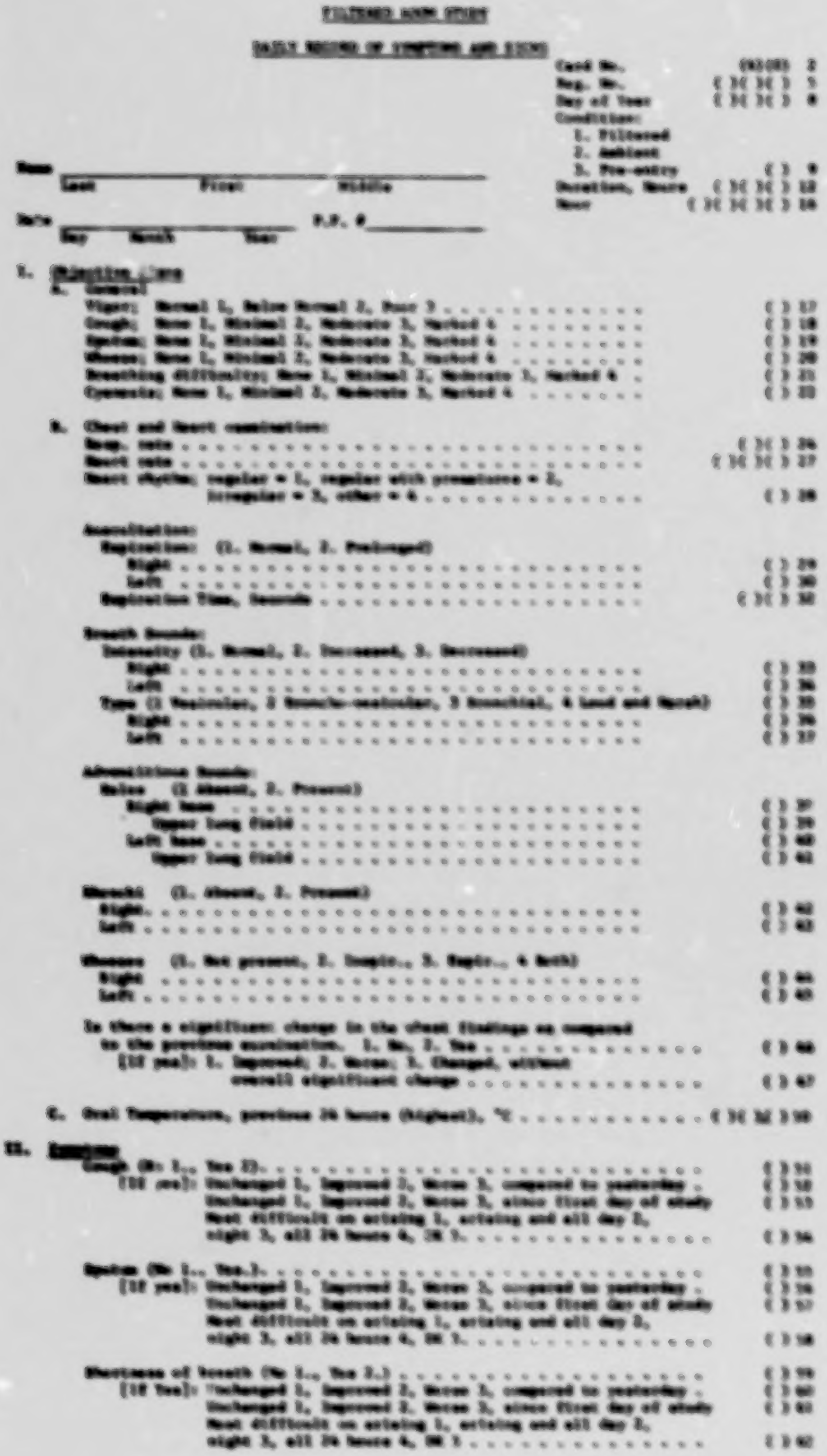
Figure 6 — Data Form: Helium Dilution

FILTERED ROOM STUDY

AIR POLLUTANTS

Card No.	(7)(1) 2	Card No.	(7)(2) 2
Room No.	( )( )( ) 5	Room No.	( )( )( ) 5
Day of Year	( )( )( ) 8	Day of Year	( )( )( ) 8
Condition:		Condition	
1. Filtered		1. Filtered	
2. Ambient		2. Ambient	
3. Pre Entry	( ) 9	3. Pre Entry	( ) 9
Interval	( )( )( ) 12	Interval	( )( )( ) 12
Time of Day	( )( )( )( ) 16	Time of Day	( )( )( )( ) 16
CO, ppm	( )( ).( ) 19	CO, ppm	( )( ).( ) 19
NO, ppm	( ).( )( ) 22	NO, ppm	( ).( )( ) 22
NO_2, ppm	( ).( )( ) 25	NO_2, ppm	( ).( )( ) 25
Oxidants, ppm	( ).( )( ) 28	Oxidants, ppm	( ).( )( ) 28
Temp., °C	( )( ).( ) 31	Temp., °C	( )( ).( ) 31
Rel. Hum., %	( )( ) 33	Rel. Hum., %	( )( ) 33
Time of Day	( )( )( )( ) 37	Time of Day	( )( )( )( ) 37
CO, ppm	( )( ).( ) 40	CO, ppm	( )( ).( ) 40
NO, ppm	( ).( )( ) 43	NO, ppm	( ).( )( ) 43
NO_2, ppm	( ).( )( ) 46	NO_2, ppm	( ).( )( ) 46
Oxidants, ppm	( ).( )( ) 49	Oxidants, ppm	( ).( )( ) 49
Temp., °C	( )( ).( ) 52	Temp., °C	( )( ).( ) 52
Rel. Hum., %	( )( ) 54	Rel. Hum., %	( )( ) 54
Time of Day	( )( )( )( ) 58	Time of Day	( )( )( )( ) 58
CO, ppm	( )( ).( ) 61	CO, ppm	( )( ).( ) 61
NO, ppm	( ).( )( ) 64	NO, ppm	( ).( )( ) 64
NO_2, ppm	( ).( )( ) 67	NO_2, ppm	( ).( )( ) 67
Oxidants, ppm	( ).( )( ) 70	Oxidants, ppm	( ).( )( ) 70
Temp., °C	( )( ).( ) 73	Temp., °C	( )( ).( ) 73
Rel. Hum., %	( )( ) 75	Rel. Hum., %	( )( ) 75
Year	( ) 80	Year	( ) 80

Figure 7 — Data Form: Air Pollutant Concentrations



FILTERED ROOM STUDY

DAILY RECORD OF SYMPTOMS AND SIGNS

Card No. (9)(0) 2
Reg. No. ( )( )( ) 5
Day of Year ( )( )( ) 8
Condition:
1. Filtered
2. Ambient
3. Pre-entry ( ) 9
Duration, Hours ( )( )( ) 12
Hour ( )( )( )( ) 16

Name ______________________________
Last First Middle

Date ____________________ F.P. #__________
Day Month Year

I. Objective Signs
A. General
Vigor; Normal 1, Below Normal 2, Poor 3 . . . . ( ) 17
Cough; None 1, Minimal 2, Moderate 3, Marked 4 . . . . ( ) 18
Sputum; None 1, Minimal 2, Moderate 3, Marked 4 . . . . ( ) 19
Wheeze; None 1, Minimal 2, Moderate 3, Marked 4 . . . . ( ) 20
Breathing difficulty; None 1, Minimal 2, Moderate 3, Marked 4 . ( ) 21
Cyanosis; None 1, Minimal 2, Moderate 3, Marked 4 . . . . ( ) 22

B. Chest and Heart examination:
Resp. rate . . . . ( )( ) 24
Heart rate . . . . ( )( )( ) 27
Heart rhythm; regular = 1, regular with premature = 2,
irregular = 3, other = 4 . . . . ( ) 28

Auscultation:
Expiration: (1. Normal, 2. Prolonged)
Right . . . . ( ) 29
Left . . . . ( ) 30
Expiration Time, Seconds . . . . ( )( ) 32

Breath Sounds:
Intensity (1. Normal, 2. Increased, 3. Decreased)
Right . . . . ( ) 33
Left . . . . ( ) 34
Type (1 Vesicular, 2 Broncho-vesicular, 3 Bronchial, 4 Loud and Harsh) ( ) 35
Right . . . . ( ) 36
Left . . . . ( ) 37

Adventitious Sounds:
Rales (1 Absent, 2. Present)
Right base . . . . ( ) 38
Upper lung field . . . . ( ) 39
Left base . . . . ( ) 40
Upper lung field . . . . ( ) 41

Rhonchi (1. Absent, 2. Present)
Right . . . . ( ) 42
Left . . . . ( ) 43

Wheezes (1. Not present, 2. Inspir., 3. Expir., 4 Both)
Right . . . . ( ) 44
Left . . . . ( ) 45

Is there a significant change in the chest findings as compared
to the previous examination. 1. No, 2. Yes . . . . ( ) 46
[If yes]: 1. Improved; 2. Worse; 3. Changed, without
overall significant change . . . . ( ) 47

C. Oral Temperature, previous 24 hours (highest), °C . . . . ( )(32)50

II. Symptoms
Cough (No 1., Yes 2). . . . . ( ) 51
[If yes]: Unchanged 1, Improved 2, Worse 3, compared to yesterday . ( ) 52
Unchanged 1, Improved 2, Worse 3, since first day of study ( ) 53
Most difficult on arising 1, arising and all day 2,
night 3, all 24 hours 4, DK 5. . . . . ( ) 54

Sputum (No 1., Yes 2.). . . . . ( ) 55
[If yes]: Unchanged 1, Improved 2, Worse 3, compared to yesterday . ( ) 56
Unchanged 1, Improved 2, Worse 3, since first day of study ( ) 57
Most difficult on arising 1, arising and all day 2,
night 3, all 24 hours 4, DK 5. . . . . ( ) 58

Shortness of breath (No 1., Yes 2.) . . . . ( ) 59
[If Yes]: Unchanged 1, Improved 2, Worse 3, compared to yesterday . ( ) 60
Unchanged 1, Improved 2, Worse 3, since first day of study ( ) 61
Most difficult on arising 1, arising and all day 2,
night 3, all 24 hours 4, DK 5 . . . . ( ) 62

Figure 8 — Data Form:

Appetite. (Normal 1, Increased 2, Decreased 3) . . . . ( ) 63
(If 2 or 3): Unchanged 1, Improved 2, Worse 3, past 24 hours. . . . ( ) 64

Chest Tightness or Congestion (No 1, Yes 2) . . . . ( ) 65
(If Yes): Unchanged 1, Improved 2, Worse 3, past 24 hours . . . . . ( ) 66

Chest pain (No 1, Yes 2, Don't know 3) . . . . ( ) 67
(If Yes): Changes with breathing, No 1, Yes 2, Don't know 3 . . . . . ( ) 68
Changes with cough No 1, Yes 2, Don't know 3 . . . . . ( ) 69
Worse with exercise No 1, Yes 2, Don't know 3 . . . . . ( ) 70
[Question: Angina? No 1, Yes 2] . . . . ( ) 71

Year . . . . ( ) 80

Card No. [illegible]
Subj. No. ( )( )( ) [illegible]
Day of Year ( )( )( ) [illegible]
Condition:
1. Filtered
2. Ambient
3. Pre-entry ( ) [illegible]
Duration, Hours ( )( )( ) [illegible]
Year ( )( )( )( ) [illegible]

Previous 24 hours only. No 1, Yes 2, Don't know 3
Sore throat . . . . ( ) [illegible]
Nasal congestion . . . . ( ) [illegible]
Abdominal discomfort or pain . . . . ( ) [illegible]
Eye irritation . . . . ( ) [illegible]
Any effect of smog on breathing . . . . ( ) [illegible]
Was smog present in the room . . . . ( ) [illegible]
Did temperature affect breathing . . . . ( ) [illegible]
(If Yes): Improved 1, Worsened 2, Don't know 3 . . . . ( ) [illegible]
Change related to elevated 1, or reduced 2, temperature . . . . . ( ) [illegible]

III. Smoking:
Number cigarettes smoked in past 24 hours . . . . ( )( )( ) [illegible]
Number pipefuls tobacco smoked past 24 hours . . . . ( )( ) [illegible]
Number cigars smoked past 24 hours . . . . ( )( ) [illegible]

8 A.M. to 8 P.M.

IV. Sputum
Volume, ml . . . . ( )( )( ) [illegible]

Color: Clear 1, White 2, Green 3, Gray 4, Yellow 5, Yellow-gray 6, Yellow-green 7, Black-brown 8, Other 9 . . . ( ) [illegible]

Blood: None 1, Streaks or flecks 2, More than streaks 3 . . . . . ( ) [illegible]

Purulence: None 1, Non-purulent 2, Purulent 3 . . . . ( ) [illegible]

Physical character: Watery 1, Viscid 2, Layered 3, Clumped 4, Other 9 . . . . ( ) [illegible]

Odor: None 1, Moderately unpleasant 2, Foul 3, Fecal 4, Other 9 . . . . ( ) [illegible]

8 P.M. to 8 A.M.

Sputum
Volume, ml . . . . ( )( )( ) [illegible]

Color: Clear 1, White 2, Green 3, Gray 4, Yellow 5, Yellow-gray 6, Yellow-green 7, Black-brown 8, Other 9 . . . ( ) [illegible]

Blood: None 1, Streaks or flecks 2, More than streaks 3 . . . . . ( ) [illegible]

Purulence: None 1, Non-purulent 2, Purulent 3 . . . . ( ) [illegible]

Physical character: Watery 1, Viscid 2, Layered 3, Clumped 4, Other 9 . . . . ( ) [illegible]

Odor: None 1, Moderately unpleasant 2, Foul 3 Fecal 4, Other 9 . . . . ( ) [illegible]

V. Medications (No 1, Yes 2)

Positive Pressure Breathing . . . . ( ) [illegible]
With Bronchodilator. . . . . ( ) [illegible]
Number times per 24 hours . . . . ( )( ) [illegible]
Bronchodilator other than with IPPB (Nebulizer, etc.) . . . . . . ( ) [illegible]
Antibiotics (Specify) ____________________________________ ( ) [illegible]
Oral other medications ([illegible], [illegible], [illegible], etc.) . . . . ( ) [illegible]
Corticosteroids . . . . ( ) [illegible]

Year . . . . ( ) 80

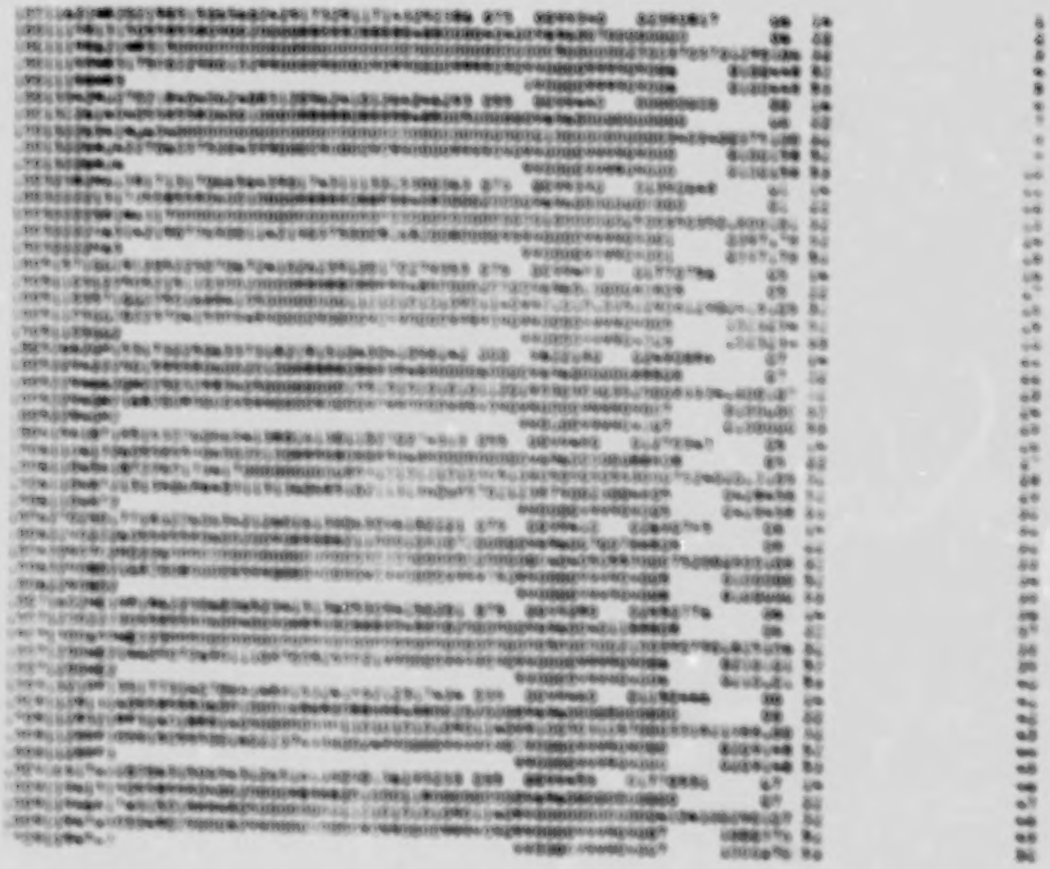


Figure 9 — Printout: Summary Form

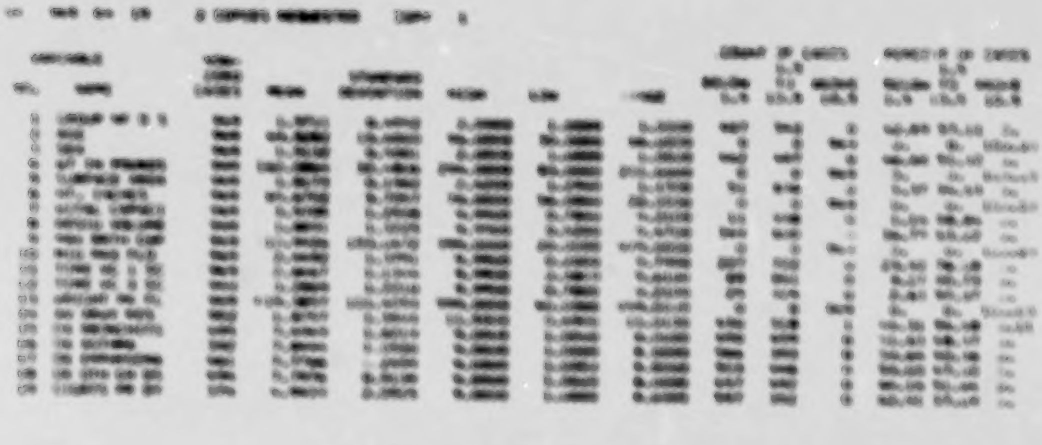


Figure 10 — Printout: Listing of Variables

GPO [illegible]

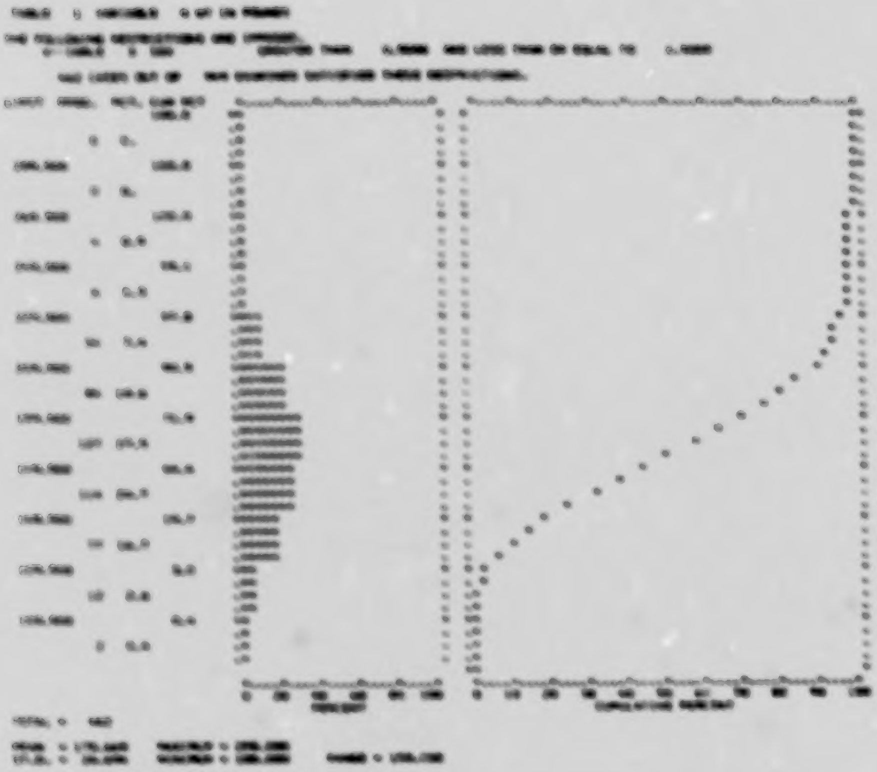


Figure 11 — Histogram and Cumulative Frequency Polygon

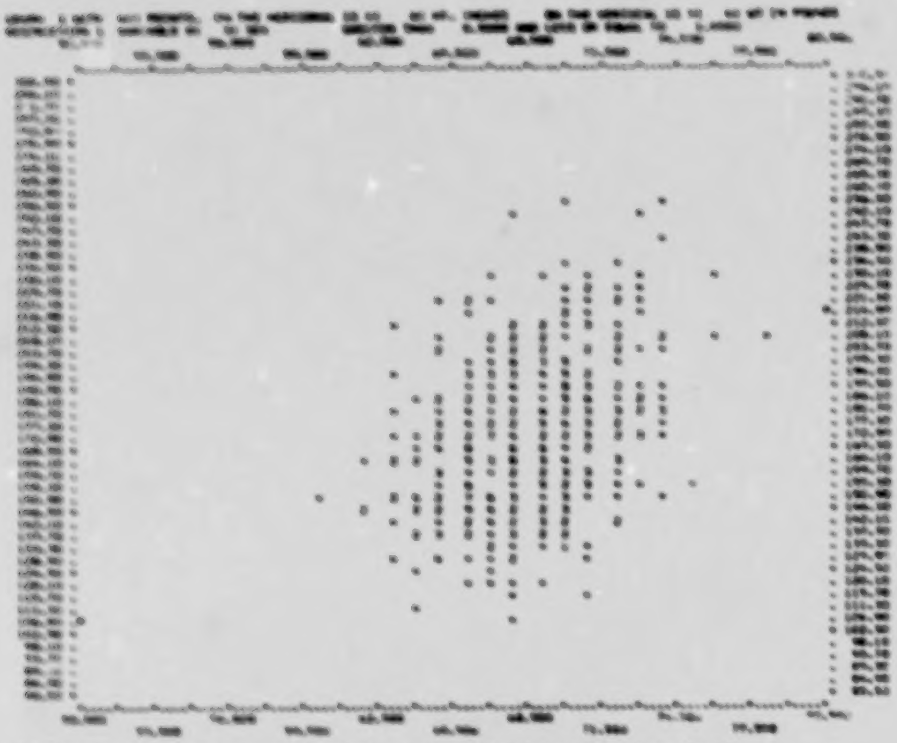


Figure 12 — Printout: Scatter Diagram

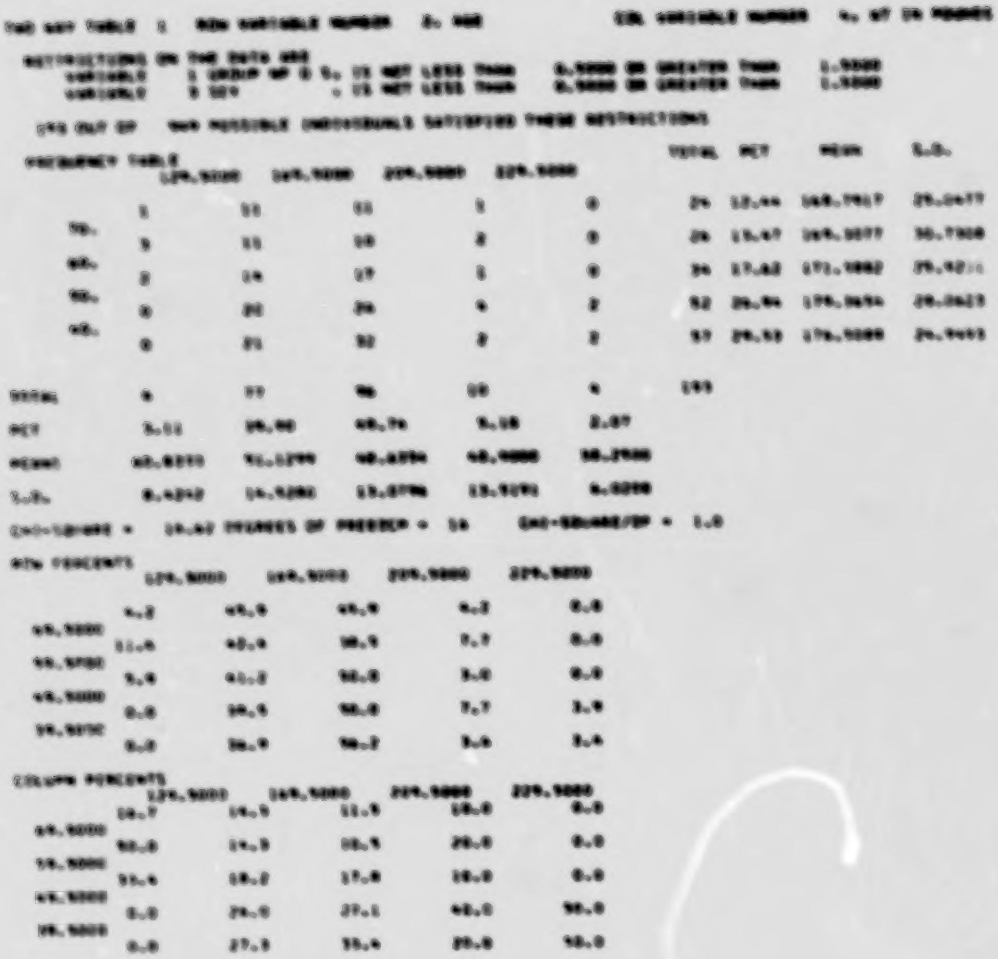


Figure 13 — Printout: Two-Way Table

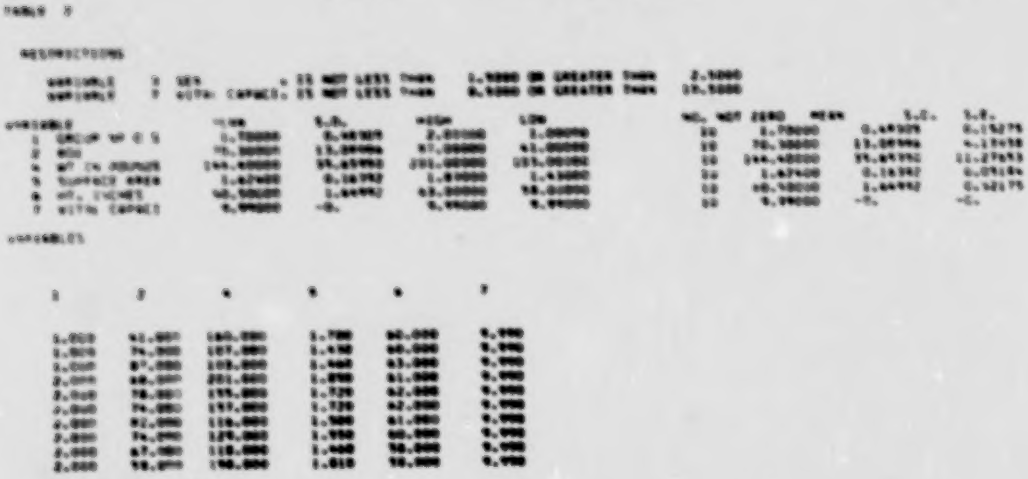


Figure 14 — Printout Showing Restrictions on Variables

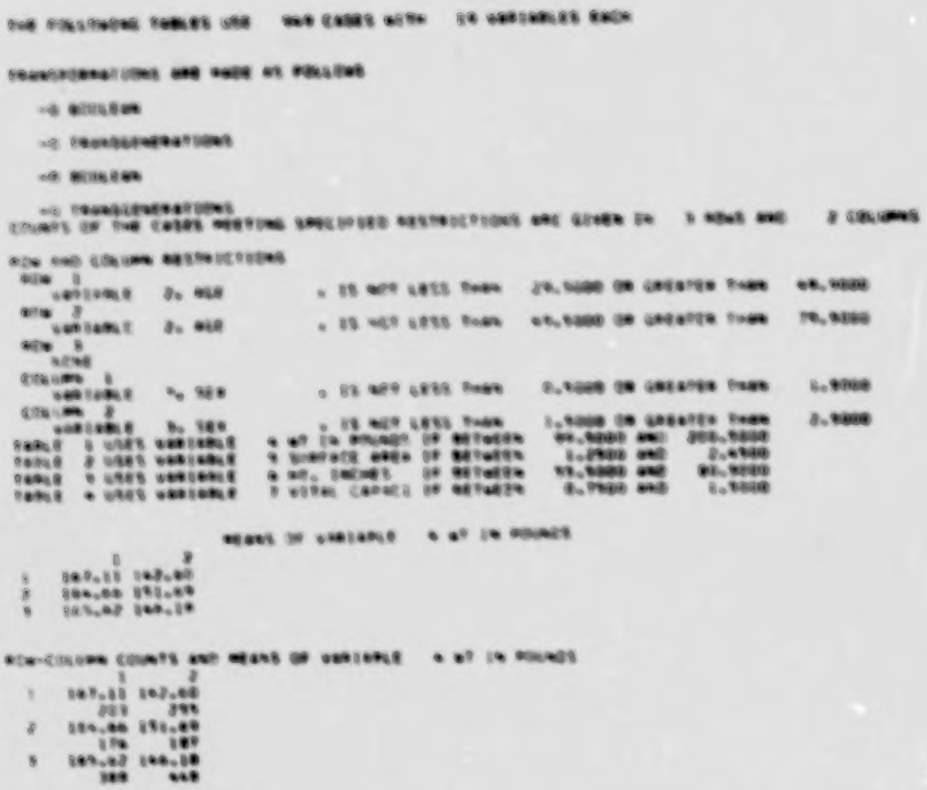


Figure 15 — Printout Showing Row and Column Restrictions

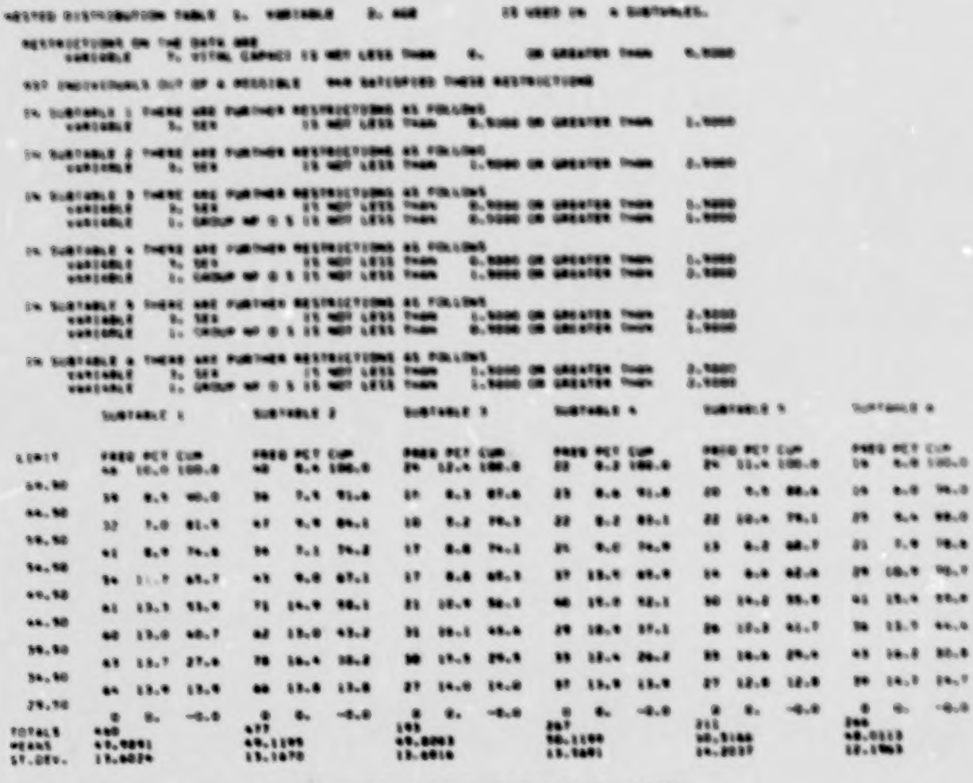


Figure 16 — Nested Distribution Table

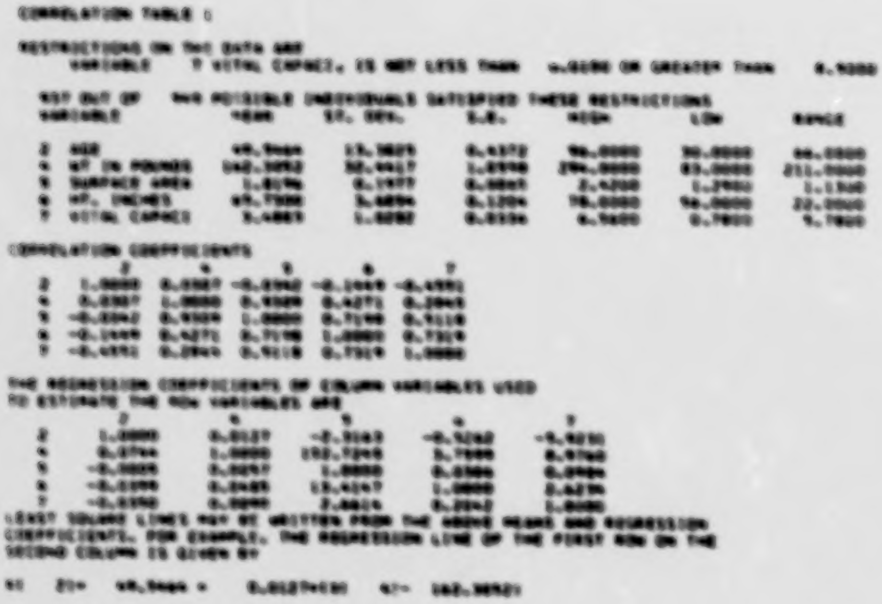


Figure 17 — Data Analysis for Selected Restrictions With Correlation and Regression Coefficients

PANEL MEMBERS

Robert Bryan
Director, Technical Services
Air Pollution Control District, Los Angeles

Dr. Paul B. MacCready, Jr.
President, Meteorology Research, Inc.
Altadena, California

Dr. Benjamin V. Branscomb
Associate Professor of Medicine
Medical College of Alabama, Birmingham

Dr. Ralph I. Larsen
Field Studies Branch, DAP
U. S. Public Health Service, Cincinnati

DISCUSSION: DATA ACQUISITION SYSTEMS

Mr. Bryan indicated that the prime question asked by an administrative group responsible for a practical data acquistion system is "Why do we measure?" This question is answered at the local level by the need to establish trends or background information. Air quality monitoring indicates whether proposed or enforced standards are being met and whether control activity is producing the desired effect.

Dr. Branscomb pointed out the inadequacy of our measuring devices in precision and accuracy. He noted that measuring devices seemed to fall into three categories: those designed principally around the chemical aspects of the measurement, those designed principally for the engineering aspects of sampling and analysis, and those oriented toward electronic interpretation of the measurement.

Dr. MacCready mentioned several items not yet considered at the meeting. He stated that the ambient air is a very poor laboratory because it varies in three dimensions and also in time. He pointed out the value of the light airplane in assessing air pollution problems. The mobile airplane appears highly flexible and relatively inexpensive in comparison with the money and manpower required by extensive ground networks. He also mentioned the use of naturally occurring topographical configurations, such as craters, which are good for special stable air mass studies. Dr. MacCready commended the use of tetroons as a means of remaining with a particular air parcel and noting its change during transport. The atmospheric laboratory can be made more quantitative by introducing more flexible means of measurement and analysis.

Dr. Branscomb noted that medical disease and its ramifications are just as difficult to define as air quality. One of the problems in studies of biological variations in subjects is a lack of definite knowledge that medical effects are due to a specific pollutant. The engineer appears to be ahead of the physician in defining variables. Diagnosis alone is unsatisfactory as a goal for measurement. Although approximately 10 percent of the adult population over 40 have emphysema, the medical body is divided in its opinion of just what comprises this disease. Most of the information accepted by the medical profession is inferential; often there are no clear-cut proofs to substantiate medical knowledge. Therefore, scientists have every right to question the physician when he submits a medical diagnosis or finding as a goal for measurement.

Dr. Branscomb pointed out the uselessness of a static tool such as the x-ray for diagnosis of emphysema. Present instrumentation used for measuring loss of respiratory function is not sensitive enough to determine small incremental amounts of such loss due to possible air pollution effects on the lungs. Commonly used instrumentation is seriously impaired by overshooting and damping effects and does not produce valid data at the frequency cycle associated with human breathing rates. A spyrometer developed at the Alabama Medical College was cited as a considerable improvement over existing instrumentation, but this unit is still barely adequate to meet the investigative needs. Dr. Branscomb noted that in the Alabama respiratory study the item that correlated best with reduced pulmonary function was positive response to the query "Does weather influence your breathing?"

Asked whether airborne particulates are important in health considerations, Dr. Branscomb replied that there is no evidence that particles alone cause detrimental health effects. He noted, however, that a recent study with guinea pigs showed that carbon particles exposed to nitrous oxide produced lesions in the lung. When the animals were subjected to nitrous oxide and carbon particles individually, they exhibited no such tissue damage. Although these two substances produce no noticeable health defects individually, a detrimental effect was caused by their inhalation simultaneously.

Dr. Branscomb further noted that emphysema has now become the second highest cause of disability in the United States; the magnitude of this problem has become so great that we must act on the basis of preliminary information. We must continue to accumulate facts regarding the effects of air pollutants in the production of emphysema and other disorders, so that industry may act in the public interest rather than being guided by public imagination. Mr. Nader commented on a particulate (sulfur dioxide) study now under way at the Harvard School of Medicine. Present indications are that particle size is important, since smaller particles seem to yield a greater physiological effect.

SESSION 5: Measurements of Water Environment

Chairman: Leo Weaver
Chief, Water Quality Section, Basic Data Branch
Division of Water Supply and Pollution Control
U. S. Public Health Service

Samuel S. Baxter
Water Commissioner and Chief Engineer
Water Department, City of Philadelphia
and
Joseph V. Radziul
Chief, Research and Development
Water Department, City of Philadelphia

SUMMARY

The Philadelphia Water Department and the U. S. Geological Survey have established a water quality monitoring network along the Delaware River estuary. The status of automation and its application to the water industry are evaluated. If it is assumed that standard biological waste treatment, low flow augmentation, and treatment of water supplies by conventional plants and methods will continue to be used to deal with pollution, the required data acquisition systems are already on the market. These include equipment for data transmission, recording, storage and retrieval, and the actuation of secondary devices. The missing elements are certain sensing and detecting devices and the full knowledge of what parameters or variables reveal the cause and effect relation within a system. These items are explored in detail, including the economics of an automatic system. The authors believe that we are at the point of no return — that automation is the key to the water industry today.

DATA ACQUISITION SYSTEMS IN WATER SUPPLY

Back in 1960, in Cincinnati, and under the same auspices as this meeting, the senior author presented a paper, "High Quality Water Without High Quality Data — Is It Possible?"[1] At that time the City of Philadelphia had just placed in operation on the Delaware River a modern water treatment plant with complete facilities for automatic chemical application and with other automatic plant operational features. This Torresdale Plant has been termed an "automatic" or "push button" plant and is considered to be one of the most modern facilities of its kind. The principal purpose of the 1960 paper was to tell about the new Load Control Center in Philadelphia, which gathers intelligence from various instruments throughout the City through a system of micro-wave stations and land wires and with this information maintains surveillance over the distribution system, logs data for record purposes, and maintains supervisory control over pumping stations. The paper also mentioned the beginning of the cooperative venture between the Philadelphia Water Department and the U. S. Geological Survey (USGS) in establishing a monitoring network for river quality measurements along the Delaware River estuary.

The 1960 paper posed challenges for water quality treatment and control, ranging from practical available instrumentation to "blue sky thinking." It was hoped from the discussions that arose in connection with the paper that the water industry and public and private research in government and industry appreciated the problems and would take real action in attempting to solve them.

As we appraise the situation today, very little has been done. The title of the paper given today is a paraphrase of the one given in 1960. Although this paper will attempt to view the problem of Data Acquisition Systems in water supply from a broad viewpoint, it will be natural that many of the illustrations will evolve around the Philadelphia plants on the Delaware and Schuylkill Rivers. Some of these illustrations may have a limited value to others, since the Delaware River source is a tidal estuary.

We look at Philadelphia's modern water treatment plants, its new Load Control Center, and the automatic monitoring stations on our rivers, and wonder where we go from here. The answer comes with a real impact to the authors of this paper. It is that we are at the point of no return, not only for ourselves in Philadelphia, but for everybody in the water industry. Automation becomes the key word. It now affects a substantial segment of our entire society, and its accelerated impetus is becoming everyone's interest and responsibility.

Automation cannot be ignored because in many places it has demonstrated that it provides more goods and services of superior or higher quality at lower costs. The report of the Committee on Public Works of the U. S. Senate[5] emphasizes the need to provide maximum service at minimum cost for all public works.

Private industry is making more and more use of automation in process industries and in manufacturing in general. The water industry and other related public works operations, generally tied into government, are far behind in developing and using automatic features.

One reason for this is the fact that each water industry is a utility, whether it is governmentally owned or privately owned. For the governmentally owned water utilities, there is the complete absence of the profit motive that provides the stimulus for private industry to lower costs and to increase quality. In the privately owned utilities, the regulation by commissions may have a somewhat similar effect.

In the water industry therefore, we should try to find a substitute for the profit motive. We should have the desire to turn out water of better quality and to reduce operating and capital costs. It would seem to the authors that in the water industry it is only through automation, instrumentation, and remote control that operating costs and reduction in personnel can be achieved. We recognize that new processes may be invented and developed, but point out certainly that these should be fully automated. The water industry is not much different from many other process industries that operate every hour of the day and week, and we should take a lesson from our brothers who, in such industries as power plants and oil refineries, have used automatic features for many years.

We cannot afford to ignore the effects automation has throughout industry, and with particular reference to skills of personnel and working hours of personnel. If the work week in industry in general will be reduced as a result of automation, the water industry will have to face this problem, including competition for skilled and professional people.

The 1960 paper suggested that the next step beyond automating treatment plants and automatic raw water sampling was a digital computer for quality control. The paper should have properly stressed the more obvious first need, which was the need to make all operations automatic so that they would eventually lead to automation.

We would then be ready for cybernation, which might be described as the science that deals with the marriage of automated systems and machines with computerized analyzing and decision-making machines. Perhaps, we may be able to strike some middle ground between cybernetics or blue sky thinking and waiting or doing little.

Blue sky thinking will entail the complete evolution of a new system of treatment. This would call for the invention of new water treatment processes based upon presently unknown concepts. The optimum system might be one in which water and sewerage systems are integrated in a realistic, functional, recycling total water use relationship.

It would be wonderful if we could have all of this. While waiting for the inspiration to bring it about, what can we do now that is in the realm of practical reality and achievement? Let us assume that standard biological treatment and low flow augmentation as we know them today will be the facilities to deal with pollution, and that conventional plants and methods with minor modifications will continue to be utilized for water treatment and quality control.

If we start from this assumption, it seems that the required Data Acquisition Systems that could possibly be used in water supply systems are available on the market today. This includes equipment for data transmission, recording, storage, and retrieval and the actuation of secondary devices.

What then is missing? The missing elements are certain sensing and detecting devices, and the full knowledge of what parameters or variables reveal the cause and effect relation within the system. This paper will explore these items in more detail from the viewpoint of the municipal water manufacturer.

SOURCE OF SUPPLY

The working agreement between the Philadelphia Water Department and USGS[6] has been centered in recent years on the establishment of a system of water quality monitoring stations along the Delaware River estuary. The initial objectives of the Water Department in this work were:

1. To maintain a surveillance network and to warn of spills above and below the raw water intake so that remedial action can be taken in time at the treatment plant.
2. To obtain a continuous record of certain water quality parameters for analysis so that some of the major cause and effect relationships within the estuary ecosystems can be resolved.
3. To provide essential raw water characteristics as input for the eventual automation of the water treatment plant, or during the transition period to provide more meaningful data for better plant control.

At the present time, the U. S. Public Health Service is engaged in a water quality—pollution abatement survey of the Delaware River estuary. This work is being done in cooperation with, among others, the states of Pennsylvania, New Jersey, and Delaware, and the Philadelphia Water Department. In this survey, greater need and use have been found for the water quality data obtained from the monitoring stations. Recently Quigley[8] cited another threefold purpose for which the data from the monitoring stations may be used in part:

"A. The determination of the cause and effect relationship between pollution from any source and the present deteriorated quality of water in the estuary.
B. Development of methods of forecasting variation of water quality due to natural and man-made causes.
C. Methods of optimal management, including necessary waste removal and flow regulation to control the quality of water in the estuary for municipal, industrial, agricultural, fisheries, recreation, and wild life propagation."

Parker[7] had considered the six parameters of pH, oxygen, conductivity, temperature, turbidity, and sunlight intensity to be of major significance for the [illegible]-sponsored automatic water quality monitoring stations on the non-tidal Delaware River above Trenton. Cleary,[8] during the development of the Ohio River Valley Water Sanitation Commission (ORSANCO) Robot Monitor, questioned and explored water quality parameters for the purpose of determining the minimum number of significant parameters that would be most useful in the Ohio River operation. Cleary and Parker were of accord with the exception of oxidation reduction potential (ORP), chlorine ion, and turbidity. Thomann's dissertation[9] on the use of systems analysis to describe the time variation of dissolved oxygen in the tidal stream says that dissolved oxygen alone is fundamentally a function of six variables.

"A. The velocity field and diffusion.

B. The temperature field.

C. The salinity field.

D. The presence of organic matter capable of utilizing oxygen in its stabilization (bio-chemical oxygen demand).

E. Photo-synthetic action by aquatic plants.

F. The presence of chemicals which would utilize or produce oxygen in certain reactions."

Thomann and Sobel[10] describe techniques for the forecasting and optimum management of water quality in an estuarine environment. These techniques are predicated upon the understanding of water quality variations. O'Connor[11] in his discussion on the oxygen balance of an estuary gave no consideration to photosynthetic oxygenation in the Delaware River. On the other hand, a study[12] on the same river by Dr. Hall of Johns Hopkins University, with the cooperation of the Philadelphia Water Department, produced real evidence that photosynthesis is a major contributor of oxygen. These items are noted in an attempt to show the nature and complexity of the unknowns and their entire relationships. As our knowledge of the Delaware estuary has increased arithmetically, our awareness of our ignorance has increased geometrically.

Reid[13] brings out the complex of interrelating factors involved in the study and use of streams. From his review of estuarine streams, there can be seen the many disciplines involved: biology, chemistry, physics, geology, hydrology, hydraulics, mathematics, oceanography. These items and others form the present day concept of the ecology on which the life and use of our streams is based.

Some of the questions for which answers are needed are

1. What significant parameters of water quality should be measured, for an alert system, for treatment plant control, for a quality forecasting system, for a river management system?
2. What should be the periodicity or time interval in collecting specific data?
3. What are the cross correlations of these parameters?
4. Are there any synergistic relationships between the parameters?
5. What is being accomplished to develop instrumentation that can gage quantitatively those essential parameters, such as BOD, that are not being measured automatically at the present time?

If in Philadelphia we could have the answers to all these questions, we could make further advances and progress in a fully automated water quality treatment operation. It would appear to the authors that all users of inland surface waters will be faced with answering these questions at some time in the future, in view of the increasing demand for water and the increasing pollution abatement problem. We must close the knowledge gap and develop the missing sensing and detecting devices before we can optimize the use of data acquisition systems for sources of water supply.

Much is being done in determining causal relations in the study now under way in the Delaware estuary. Without the continuous-type data that are made available only through the use of instrumentation from the new monitoring stations, progress on the study would be much slower. Further use of these stations will be made, when the Public Health Service installs on each monitoring station a digital recording system that will handle up to 10 variables. There is some thought on our part that we may eventually need 15 or 20 positions.

THE WATER PURIFICATION PLANT

Although it is apparent that the United States is in the midst of a science and technology revolution, with many new products and procedures having come into existence only since World War II, no new major concepts have been recently developed in the water treatment field. A Robert A. Taft Sanitary Engineering Center report[4] notes that "the basic methods of municipal water treatment have not changed substantially for almost 50 years." Erdei[2] points out some minor advances in water treatment, while spot-lighting the urgency that "in the era of scientific hygiene, the specifications for water, in accordance with the physiologic needs of man, must be more meaningful and exact."

Busch[16] reports that Dr. Keilin of Aero Jet General Corporation invented a new thin plastic membrane that filters salts and bacteria of body wastes, and viruses and detergents, and that "this filter could save 90% of the water that cities now discard." Without discounting such new processes unduly, we believe that the conventional plant and methods will be with us and in use for some time. What is said following is based on this assumption, since the authors believe that progress in automation and instrumentation should be made now.

INSTRUMENTATION AND AUTOMATION IN THE WATER PURIFICATION PLANT

Although the experience of the authors has been primarily with large water treatment plants located on large rivers, we believe that the comments and suggestions that follow may also apply to smaller operations. No one should write off instrumentation and automation simply because of size.

Any consideration of instrumentation and automation of water purification processes quickly encounters the obstacle that in two of the most important areas where control is needed — coagulation, and taste and odor removal — no signal is available on which to hang instrumentation. There is no analytical method for directly indicating what steps are to be taken to produce coagulation of a specific water or that proper coagulation has taken place. Similarly, no signal is available to indicate that compounds are present in water that must be processed to remove bad taste and odor or that the final processed water is free of taste and odor objections.

Instruments are available for continuous determination of turbidity, dissolved oxygen (DO), color, temperature, radioactivity, pH, ORP, specific conductivity, phenols, residual chlorine, and chlorine demand. In general, these instruments are reliable and are reasonably economical in cost of operation; they give greater accuracy and permit greater frequency of testing than manual means and techniques, and they have ability to actuate secondary devices.[11]

Since most purification operating problems are due to fluctuations in raw water characteristics and since any of the above instruments will indicate successfully the fluctuations in the parameter it measures (and record the same), it is surprising that only in the case of chlorine residual and demand instruments is there direct, general application of these instruments to control of the water purification process.

There may be some use of the turbidity instrument in adjusting the coagulant dose, but since there appears to be no direct relationship between this variable and the amount of coagulant, the turbidity analyzer cannot be converted into a control mechanism. The automatic pH recorder is of some help, but its use as a direct-control instrument is limited at the present time.

This leaves the two key purification processes of coagulation and taste and odor control without a method of direct measurement or means of controlling procedures with instruments.

There is apparently the lack of complete information about coagulation processes and the chemicals used in these processes.[12] Without such knowledge, the plant operator can do no better than employ trial and error empirical methods. Although a good operator can obtain good results most of the time, such methods can result in waste or misuse of chemicals and poor results. To be on the safe side, many operators probably overdose with coagulants.

To produce an acceptable water from the standpoint of taste and odor, plant operators must rely on a periodical manual performance of a time-consuming test, both of the raw water and water in process. At this point, the control chemist is faced with the fact that sensitivity to taste and odor in water varies greatly with individuals. It is possible that in many plants some of the duty laboratory men are not capable of performing an acceptable odor control test.

There are several reasons for the limited use of continuous automatic-analysis instrumentation in water purification plants. Some of these are the absence of equipment to measure some of the most important factors directly; the fairly high cost of instruments available; the belief that instruments are so complicated that only a highly trained man can keep them in satisfactory operation; and a lack of full knowledge of the benefits to be gained by continuous sampling and analysis.

The automatic residual chlorine instrument is an example of these points, since it is in fairly common use, is reliable, and can be used to control an essential water purification chemical. The cost of these instruments at the present time probably eliminates them from consideration by managers of small or medium-size plants. On the other hand, continuous automatic analysis of residual chlorine content at several points in the purification process would seem to be a valuable tool. If more plants would use them, the price would probably come down.

Instrumentation has not yet caught up with the basic requirements of water purification plants, and part of this may be due to the lack of fundamental knowledge of water purification processes and controls. This is a matter that affects everyone who uses public water supplies. It points out the need for research on a national basis.

This need for research on a national basis should not, however, prevent individual research by operators of water purification plants. In the Philadelphia Water Department, its water quality and research divisions conduct studies whose objective is the development of better methods of control of chemical dosages and greater efficiency in the use of chemicals. Bean, Campbell, and Anspach[17] have made studies of the Zeta Potential method of coagulation control and efficiency at the Torresdale Water Treatment Plant. Studies have also been made on the use of polyelectrolytes[18] in coagulation. One conclusion that has been reached is that better control of chemical treatment is possible after the establishment of optimum rates by measurement of turbidity and Zeta Potential. Our opinion is that the door could be open for complete automatic control of coagulant application if instrumentation could be developed for measuring Zeta Potential automatically and continuously.

Here is the point where science and research meet head on with economics. The Philadelphia Water Department spends about $1,150,000 annually for chemicals used in this water treatment process, including a large amount of alum. If automatic monitoring and controls could save 5 percent of this, the annual saving of $57,500 would carry the capital charges on a large amount of instrumentation. This is the carrot we would like to dangle before the noses of both water treatment operators and instrument manufacturers. If the 5 percent seems too high, even 3 percent would do wonderful things.

PLANT DESIGN

In the continuing process of evolution in all of our manufacturing and commercial operations, there runs the trend of a minimum of attention and physical effort by plant operators. This is an element the water treatment plant designer and operator should not ignore, if for no other reason than the difficulty in attracting and holding competent personnel.

For this reason, serious consideration should be given in new plants for provision of centralized control and automatic operation. If fully automated operation is not possible at the present time, the possibility of such operation in the future should not be ignored.

DISTRIBUTION SYSTEM

Considerable progress has been obtained in automating various distribution elements of a water supply system. The Load Control Center at Philadelphia[20] now transmits and records to a central point full information about an entire system and also controls from this same central point nearly all of the pumping stations. The aim is for more complete and economic control of the system, however, since yearly power costs are $1,600,000 and a small saving in this amount would justify additional instrumentation.

Because of the ability to serve any distribution district with two or more pumping stations with different power factors and pump efficiencies, a fairly complex problem is posed when the most economical combination of dispatching water to a district is considered. It seems to the authors that we will only realize minimum power expenditures when load dispatching is regulated by computers with complete automatic equipment. Brock[19] has reported experience in the Dallas City water works of developing a computer program for distribution network operation.

Bean[3] stated that there is great need to obtain analytical information on water at the point of delivery to the customer. In this case, we are confronted again with water quality criteria for which there may not be a sensor. Of primary interest to the customer would be clarity, palatability, tastes and odors, and pressure. The treatment operator would also be interested in these factors, plus basic chemical and bacteriological criteria.

PERSONNEL

There is another and rather odd form of data acquisition system in the presence of homo sapiens. It has been said in several places that the half-life of the present engineering graduate is 10 years, unless he updates himself. If the water industry is to proceed along the path of automation, being followed by other industries, it will have to update its personnel or its homo sapiens data acquisition system. It seems mandatory and imperative that industry and education accept the responsibility of updating the men in this field. There are many ways of doing this, but it should be thorough and complete. Without belittling in-training courses or self-teaching, there is also room for the 6-month or 1-year sabbatical for instruction of personnel who can assimilate the new techniques. The water industry itself will have to recognize that the cost of such complete training is as important as the cost of a new building or of power or chemicals. Rather it should be said that it is more important than these items.

The water industry has talked about the shortage of trained men, ranging from top professionals to technicians and operating personnel. A fair share of the best men must be attracted to the industry. They will come if they are given the same challenges and the same modern operating features they will find in other industries.

Much of water treatment plant operation, including the laboratory work, is routine. Good men will shy away from work that is all routine. They need the challenge of some amount of research and application of new ideas. The field of automation in water purification can provide that.

ECONOMICS

As indicated earlier, economics and cost cannot be ignored in the field of automation and instrumentation. A water works manager can only justify the cost of instrumentation if he can prove that it will result in lower operating costs or in better quality, for which an economic value can be given. This was outlined in detail in the discussion of the Philadelphia Load Control System.[20] Therefore, instrumentation in the treatment plant must be balanced by a reduction in cost of personnel, in chemicals, and in general maintenance and operation.

The other factor that cannot be ignored is the cost of instrumentation. If development costs must be reclaimed through the sale of only a few instruments, the cost of these instruments will be high. If there is much use and demand, the costs will be lower. Since instrument cost is a basic factor in this, possibly it behooves all of us in the water industry to lift ourselves by our boot straps by using new instrumentation wherever possible.

REFERENCES

1. "Water Quality Measurement and Instrumentation" — Transactions of Seminar, R. A. Taft Sanitary Engineering Center, Cincinnati, Ohio (August 1960).

2. Erdei, Joseph F. "Advances in Water Treatment" Journal AWWA 55:845 (July 1963).

3. Bean, E. L. "Progress Report on Water Quality Criteria". Journal AWWA, 54:1343 (November 1962).

4. "Biological Problems in Water Pollution" — Transactions of Seminar, R. A Taft Sanitary Engineering Center, Cincinnati, Ohio (April 1959).

5. Committee Print No. 3, "Study and Investigations of Use of Materials and New Designs and Methods in Public Works" Committee on Public Works, U. S. Senate (1962).

6. Cleary, E. J., "Development of a Robot System". Journal AWWA 50:1219 (September 1958).

7. Parker, B. W., Freeberg, J. A. and Barber, S. B., "Automatic System for Monitoring Water Quality". J l. Sanitary Engineering Division, A.S.C.E. Paper 2554, SA 4. Vol. 86, p. 25, July 1960.

8. Quigley, James M., "Statement on Water Quality Management of Delaware Estuary". Presented before Natural Resources and Power Subcommittee on Government Operations, at Trenton, N. J., August 9, 1963.

9. Thomann, Robert V., "The Use of Systems Analysis to Describe the Time Variations of Dissolved Oxygen in a Tidal Stream." A dissertation in the Department of Meteorology and Oceanography submitted to the faculty of the Graduate School of Arts and Science in partial fulfillment of the requirements for the Degree of Philosophy at New York University, N. Y. (November 1962).

10. Thomann, R. V. and Sobel, M. J., "Estuarine Water Quality Management and Forecasting". Presented at ASCE Water Resources Conference, Milwaukee, Wisconsin. (May 15, 1963).

11. Jones, R. H. and Joyce, R. J., "Instrumentation for Continuous Analysis", Journal AWWA 53:713 (June 1961).

12. Larson, P. E., "Research, Needs, Priorities, and Information Services", Journal AWWA 54:657 (June 1962).

13. O'Connor, D. J., "Oxygen Balance of an Estuary", ASCE 1961.

14. Hull, C. H. J., "Photosynthetic Oxygenation of a Polluted Estuary", Report No. XIII, Low-Flow Augmentation Project, The Johns Hopkins University, January 1962.

15. Reid, W. C., "Ecology of Inland Waters and Estuaries", Reinhold Publishing Corporation, 1961.

16. Busch, H., "Pollution Problem Gets More Attention", The Ensign, July-August 1963.

17. Bean, Campbell and Anspach, "Some Aspects of Zeta Potential", Presented before Pennsylvania Section of AWWA, June 5, 1963.

18. Campbell, S. J., "Coagulation Studies — Nalco 614, Jaguar W.P.B.* and Narvon Activated Clay — Z^3". Unpublished report, Philadelphia Water Department, July 26, 1963.

19. Brock, D. A., "Closed-Loop Automatic Control of Water System Operations", Journal AWWA 55:467, April 1963.

20. Baxter, S. S. and Appleyard, V. A. — "Centralized Load and Quality Control Systems at Philadelphia". Journal AWWA 54:1181 — October 1962.

DISCUSSION

Mr. Baxter was asked whether prizes have been used as an incentive to manufacturers to develop instruments needed for environmental measurements. He indicated that although this has been considered, the main incentive for industry probably is more business and a wider market. Instrument manufacturers must have a reasonable prospect for the sale of the instruments they develop. In the meantime we may have to make greater use of the instruments now available.

Although no automatic instrumentation was in use in the Delaware River project 3 years ago, an instrument that measures six parameters continuously is now in operation. Manufacturers of this instrument anticipated a sufficient market to warrant the development costs. Mr. Baxter challenged water researchers to use more instruments in the production of water and suggested that if more instruments were purchased, the manufacturers would do more to develop cheaper and better instrumentation.

Mr. Mentink asked what use is made of the water quality data collected by the monitoring installation above the Philadelphia water intake. Mr. Baxter replied that the information from the automatic instruments is not being used to change or to regulate day-to-day operation of the water plant. If continuous valid water quality data were available, it could be used to adjust the automatic facilities used in operation of the plant.

Mr. Mentink commented that there are apparently cost differences in processing water with differences in turbidity, pH, dissolved solids, chlorides, etc., and that some savings in cost should result from knowing these different qualities. Mr. Baxter agreed that cost differences, especially with turbidity, could be significant. The cost of alum, one of the largest chemical costs in water processing, could be reduced if water quality were known in more detail. For example, if a plant operator wants to make sure that the chemicals are adequate to accomplish coagulation, he will generally use an excess. Better information on the quantity of chemicals needed would allow savings of some of this excess. As another example, if continuous information showed the presence of chemicals that cause taste and odor, such as phenol, an operator could initiate the addition of activated carbon, which would not ordinarily be used.

Paul De Falco, Jr.
Director, Raritan Bay Studies Project
Division of Water Supply and Pollution Control
U. S. Public Health Service

SUMMARY

In a data acquisition system for water quality control, the "need to know" and the ability to use the information collected must be carefully examined. Past data should be reviewed to establish the frequency and location of representative stations. The data handling and analysis system must be within the program's limitations of need, interpretation, personnel, equipment, time, and money. Data should be collected in accordance with a plan that best meets all these requirements. At regular intervals during the program of studies, data collected to date should be reviewed so that it can be determined whether needs are being met and whether alterations in the system are necessary. Any changes should be fed back in such a way that an operating system will continue to meet the program's requirements.

DATA ACQUISITION SYSTEMS IN WATER QUALITY CONTROL

An impressive assortment of data collection systems was described at the "Symposium on Water Quality Measurement and Instrumentation" held here in August of 1960. I left the meeting with the feeling that collecting data was the easiest part of the problem in our business. What I had not heard discussed at any great length was just why we were collecting these data and what we were supposed to do with them once collected. Since that time, a lot of water — I don't know what the quality was — has passed under the bridge. We are still collecting data.

Collecting data is usually a simple procedure. What complicates the picture is answering the questions:

Why are the data needed?

Where should they be collected?

When should they be obtained?

What will be done with the information?

In the development of measurement systems in water quality control we have reached the point where we must critically answer these questions before embarking upon a course of studies requiring collection of data. Data collection is expensive. It costs between $25 and $50 to analyze a sample and almost again that much to collect it. When the costs of processing and storing data are also considered, it becomes obvious that no data should be accumulated without a positive justification.

Recently, I was asked by a representative of a state water pollution control program to assist in the development of a water quality surveillance system for his state. My answer was to ask him a series of questions:

What were his agency's needs for information?
What objectives would this system be required to meet?
What parameters were of value?

What past records of water quality were available to indicate frequency of sampling and the location of stations to best describe changes in quality?

What data utilization system was available in the department to handle, analyze, and interpret the data collected?

Let us face the facts. The data acquisition system we usually adopt is a compromise between the system we need and the system we would like. We sometimes collect data because it is convenient . . . it is nice to know . . . it may be important some day . . . and because everybody else collects it. We should collect it because we have a use for it, and more importantly, because when interpreted it will answer the needs of our organization. The limits of any single element in the chain from collection to utilization should set the limits for the individual steps. The extreme shortage of skilled engineering and scientific personnel in our field does not allow us the luxury of collecting interesting or unduly refined data unnecessary to the needs of our organization.

ESTABLISH PROGRAM NEEDS

The design of a water quality control data acquisition system must evolve from the needs for the information in the operation of specific programs. These needs can generally be classified as follows:

. . . To determine compliance with a given standard.

. . . To determine or forecast the effect of a water resource project.

. . . To determine treatment needs in the use of the water.

. . . To provide water quality control.

To determine compliance with a given standard, or criterion, several levels of sophistication are available. These range from the simple go no-go decision in a measurement that indicates whether a level is being exceeded or not to the more elaborate models with built-in warning systems to indicate in advance that remedial action may be required at a future time.

Water resource projects must be evaluated for the effect they may have on the quality as well as the quantity regimen of the stream. This requires the measurement of changes by "before and after" studies of the project. Forecasting changes without knowledge of the "after" conditions is possible, although difficult. It requires careful correlation to changes encountered in similar projects elsewhere.

The measurements required in the utilization of water, such as in a water treatment plant, usually are well identified as those elements that may be controlled by treatment or that may affect the safe usage of the water.

The measurements needed to provide quality control, either through stream or waste flow regulation, are usually those that are indicative of the problem being controlled. These measurements generally are required only for the period during which the flow is available for regulation.

DEFINE OBJECTIVES

The objective of a measurement and data acquisition system is to temporally and spacially characterize the quality of the stream or body of water with respect to the parameter chosen. Definition of changes that occur in quality between periods of time or given locations satisfies the needs of most water quality management agencies; however, the degree of sensitivity to change that is required by each agency depends upon the ultimate use that the agency makes of its collected data. This in turn determines the degree of sensitivity needed in the measurement system.

Adequate time must be given to the planning of a data acquisition system. Such planning includes programs for data analysis and interpretation prior to and concurrent with data collection, as well as after collection. One must continually examine the data being collected to ensure that the objectives of characterization of base quality and the changes in quality are being met.

SELECT PARAMETERS

The decisions on the parameters to be measured are most important. If the agency's need for water quality data has been carefully defined, the parameters that are a direct or indirect measurement of the water quality need to be fulfilled are normally obvious. The problem lies in the addition of parameters that are not necessary to the needs of the agency, or to the neglect of parameters that are interrelated with the parameter desired.

The addition of parameters to a study should be carefully weighed with respect to the cost of collection and analysis as well as interpretation. Any extra cost might better be used for expanding the temporal or spacial network for sampling the parameters of direct interest. For example, an agency conducting a program to determine compliance with a bacteriological standard should weigh carefully the productivity of additional tests for chemical quality, as opposed to productivity of increased bacteriological examinations in the fulfillment of its mission.

Conversely, the interrelationship of some parameters requires the measurement of additional parameters so that the phenomena being observed can be described more adequately. An example of this is dissolved oxygen, which is interrelated with temperature, conductivity, BOD, turbidity, algae, solar radiation, wind, and hydraulic characteristics.

The parameters themselves often dictate limitations on the sampling system. Many of the parameters we are most concerned with cannot presently be measured automatically. Others require such extensive laboratory work that the temporal or spacial grid desired for proper interpretation is limited. Modification of test methods is sometimes warranted to establish screening procedures. When the presence of the parameter has been qualitatively established, a more intricate quantitative analysis can be set up. Tests for phenols are a good example.

SEARCH HISTORICAL DATA

Prior to the establishment of a data acquisition system, existing data should be carefully reviewed. Many agencies and organizations collect water quality data for special purposes. Much of this information, although not necessarily oriented to the need of the proposed data acquisition system, provides background knowledge of the type of variability existent in waters to be monitored. Information on the type of temporal or spacial variability is a prerequisite to the development of a good measurement system.

Sources for information of this type include past studies by the agency concerned; by other public agencies in the water quality management field; by public, private, and industrial water suppliers; and by sewage and industrial waste treatment plants. The sampling and analytical procedures used by others must be carefully evaluated to qualify the value of the data collected. On the Delaware River, we found that more than 40 groups were collecting water quality data. The analysis of these data permitted a significant reduction in the field work required to meet the program's objectives.

DEVELOP DATA UTILIZATION

At this point in the development of a data acquisition system, it is necessary to develop more completely the data utilization program to be followed. Included in this program, in order of consideration, are data interpretation, data anlysis, and data handling.

Data interpretation is simply the meaning that is to be placed upon the possible values that may occur in the selected parameters. By this time, the specific questions to be answered by the study should have been formulated. These questions formulated in terms of hypotheses to be tested by measurement enable the statistical consultant to design an adequate program of sampling, data handling, and data analysis.

Data analysis may be limited by the resources of the agency. The collection scheme must, however, meet the limitations of the analysis system available to the agency. And, in turn, the data handling system, whether it be simple forms, or punch cards, or tape, must mesh with the data analysis system available.

It is valuable to test the system chosen with historical data to determine whether it responds with the required degree of refinement. Additional tests of the system should be made at regular intervals during the study to determine whether the collection system is meeting the needs of the program. The feedback from this series of checks should augment or correct the collection program as needs are determined.

DETERMINE FREQUENCY OF MEASUREMENT

Another area of decision confronting the engineer planning a data acquisition system is the frequency of measurement This, again, is determined by the "need to know." At times a continuous measurement of quality is required — such as an alarm or alert system in which a parameter goes out of control and immediately requires remedial action. The water treatment plant is an ideal example of this area of decision. In other cases, a much longer time interval between measurements meets the needs of the agency.

In most cases, the changing quality of a body of water is being characterized and for this a knowledge of the time behavior of the chosen parameter is required. Most parameters vary with time, i.e., with natural changes in water flow and temperature. Superimposed upon these are the transient effects of waste discharges. The design of a measurement system requires a knowledge of the types of changes that occur and the periodicity or trend of these changes. The data collection program must be designed statistically to develop meaningful information with respect to these changes and their causes.

LOCATE SAMPLING STATIONS

The geographic or spacial distribution of sampling stations requires the same careful consideration that has been given to the other elements of the system. Where possible, continuation of existing locations should be considered to give continuity with past studies; however, the choice of stations must first meet the needs of the program. Location must be representative of the water body being sampled and indicative of the changes that are occurring in the parameters being monitored.

Consideration should be given to obvious factors such as horizontal and vertical stratification due to temperature or specific gravity. Knowledge of the location of discharges should be considered before selection of stations. A series of dispersion studies to determine proper sampling locations may be required before a final decision is made.

DISCUSSION

Mr. Fry commented that selection of parameters should be greatly emphasized. It may be a waste of time to collect data on a given parameter, such as BOD, simply because our profession has accepted it as an essential parameter in the measurement of water quality. Rather than obtaining simple measurements of BOD in a stream, we might attempt to measure the rate of change of oxygen consumption or perhaps the total carbon content of a water. We should look for direct measures of water quality rather than trying to apply formulas and empirical approaches in our attempts to understand streams by indirect measures.

Mr. Radziul observed that the first two papers presented had definitely carried the undertone that we do not know enough about water quality and that we must determine the cause and effect relationships that exist. He agreed that the yet-to-be-discovered parameters may be the controlling ones and that old ones may have to be discarded.

Mr. Weaver commented that this discussion underscores the point that regardless of the black boxes and transducers that may be developed, man—presumably with some professional background and judgment—will still be very much in the picture.

Dr. Williams pointed out the need for better communications between scientists and engineers to make effective use of the available information. For example, although the development of chemical and physical instrumentation is urgent and desirable, planktonic organisms can be used now to provide valuable information on water quality. The species diversity of the planktonic organisms in raw water is a new parameter that has been worked out and is available for use.

Mr. Stern mentioned that an instrument system has been developed that will measure total carbon in a given water or waste water sample. Experimental units are in operation now, and further development is under way. This system will be available soon, and more such instruments should become available as time goes on.

Mr. Cohen noted that the present data acquisition systems are mainly based on probe-type devices — the DO probe, pH probe, and probes for temperature, specific conductivity, and chlorides. He pointed out that investigators in air pollution are now using four or five different wet-chemical constant-feed devices that yield continuous data, and that water researchers should give greater consideration to this type of device. While the work these probes are doing for us is fine, we should be looking toward other types of devices as well.

W. L. Isherwood
Hydraulic Engineer
Water Resources Division
U. S. Geological Survey, Washington, D.C.

SUMMARY

The U. S. Geological Survey has put a great deal of effort into the development of equipment and techniques for recording river gage heights in the field in such a way that the data can be efficiently processed into river flow data by use of a digital computer. Some false starts were made, but they now have a workable system, described here, by which records for nearly a thousand river gaging stations are being routinely processed through a computer. They plan to expand the system at the rate of about a thousand stations a year until nearly full conversion of their stream gaging network is reached. In addition, they are beginning to apply similar techniques with slightly modified equipment to the recording and processing of other types of hydrologic data, such as precipitation, temperature, chemical quality of surface water, and depths to ground water in wells.

DATA ACQUISITION SYSTEMS IN HYDROLOGY

BACKGROUND

For many years the Geological Survey has collected records of river gage heights and other types of hydrologic data on strip charts. Therefore, it was natural that we should first think of automation in terms of beginning our processing by an automatic reading of the data directly from the strip charts. We spent considerable time and money trying to develop a photoelectric chart scanner, and at times it seemed that we almost had the problem licked. But we were never quite able to overcome the basic fact that we could not satisfactorily control the quality of inked lines drawn automatically on strip charts at thousands of isolated field installations. Some lines would always be too watery, other would soak into the paper or smear, and there would often be some dirt spots and occasionally paper flaws indistinguishable from the real data line. Somewhat reluctantly, we finally came to the conclusion that automatic chart reading was impractical and that we must have a really unambiguous record if we were to be able to process large masses of data automatically with any real efficiency. Punched paper tape seemed to be the best recording medium to fit our requirements, because each bit of data is represented by either a hole or no hole in the tape, with no possible intermediate condition. We then investigated several different kinds of paper tape punching devices to try to find one capable of battery operation over extended periods at isolated field installations and still simple enough so that it would not be unduly expensive. A device manufactured by the Fischer and Porter Company seemed to have possibilities for adaptation to our needs. The unique feature of this device was a system of recording on punch tape by positioning a disc containing ridges and valleys in such a combination that a complete reading in parallel mode could be punched on a wide tape with a single stroke. That is, the input could be continuously positioning the code disc until a reading was called for and then a single throw of a punching device would punch all the digits in that reading at once in a single row of holes.

DEVELOPMENT OF THE DIGITAL RECORDER

Of course, many modifications were necessary to make this existing device fill our particular needs. For instance, the original device was operated by a-c power but we needed battery operation. This took only a simple modification. The original device used a single large code disc divided into 1000 code divisions so it could record only a three-digit number, but we needed to record four-digit numbers. This required a more fundamental modification. To make it punch four digits, it was necessary to change from one large code disc to two smaller discs, each divided into 100 code divisions. The two discs were connected by a 100 to 1 worm gear so that one revolution of the low-order disc turned the high-order disc one division. Then it was necessary to devise a mechanical non-ambiguity system to prevent trying to punch somewhere between two divisions on the high-order disc. A cam and lever system was devised to adjust the high-order disc exactly to the proper discrete digit position just prior to the moment a reading is punched out. Figure 1 shows the digital recorder as finally developed.

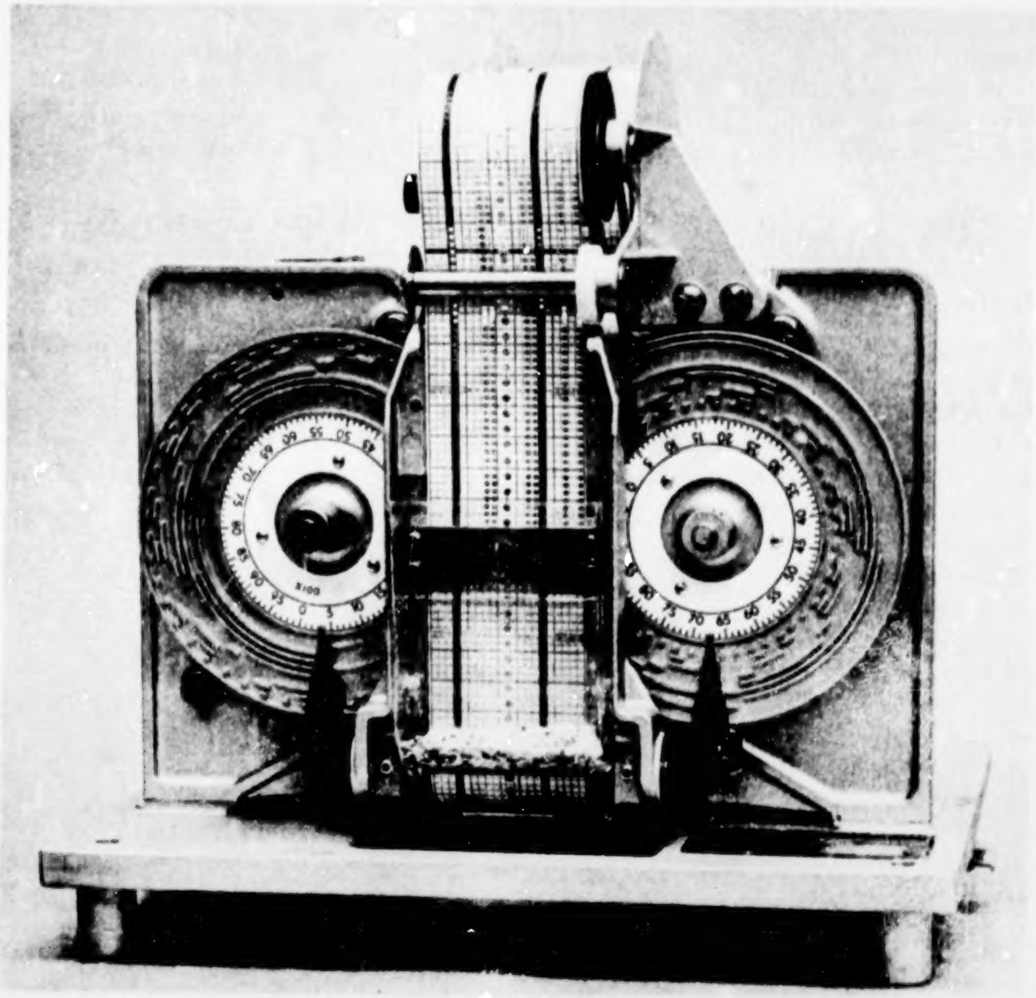


Figure 1 — Digital Recorder Developed to Record River Gage Heights.

TAPE CODING AND TRANSLATION

The format of the punched output was carefully considered. Parallel punching provided information with the least power consumption and with the simplest field instrumentation. But it was recognized that such a format would require translation before entry into any computer because computers require serial input. The necessity for intermediate translation was something of a blessing in disguise, however, because it made our field instrument entirely independent of computer requirements. We expect eventually to have several thousands of these field instruments and we expect them to last many years before replacement. But we were already in the midst of changing computers while this development was taking place and we knew that computers were changing so rapidly that we would have no way of knowing what kind of computer we might have 10 years from then. The answer seemed to be to make a translator that would have a fixed input for field tapes but would have a completely flexible output that could be easily changed to fit the input requirements of any computer we might have. This is the system we adopted. The translators we now use can punch out serial-coded paper tape in any sequence or grouping of characters required by a computer.

FIELD TESTING

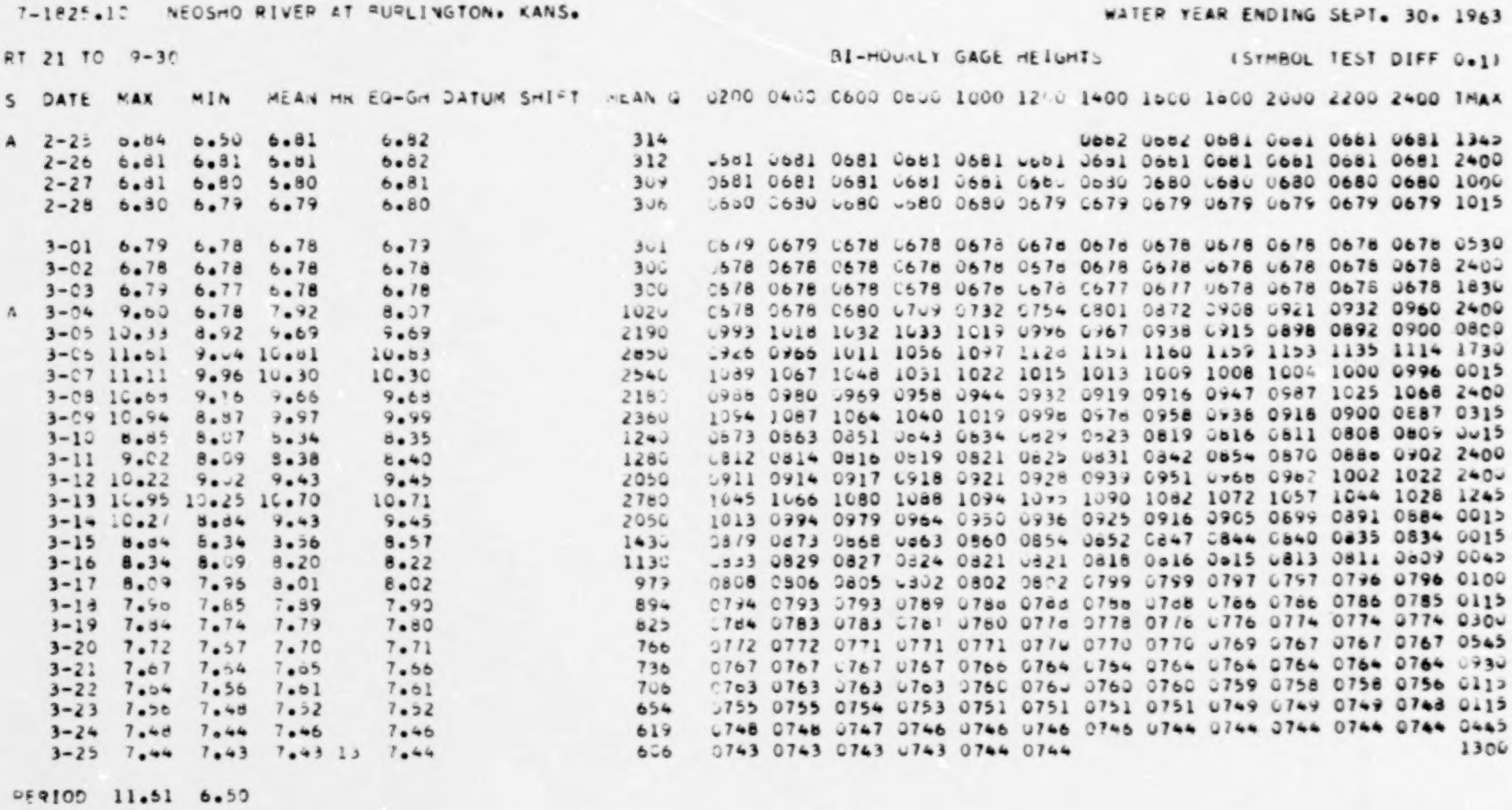
After the successful production of the basic recorder we conducted large-scale field tests covering widely varying climatic conditions and a large range of river regime types. The initial field tests involved installation at 80 gaging stations for a full year, 20 in each of 4 areas, the New England states, Alabama, Kansas, and California. After working out the few bugs that showed up in these field tests, we were really ready to start using the digital recorders in large numbers.

PROCESSING PROCEDURES

The processing done on these river flow records involves first the paper tape translation using off-line equipment, then a primary computation performed immediately on receipt of the record from the field, and at the end of each water year an updating process and a final print for publication. The primary computation computes figures of daily mean discharge plus several other useful items for each day and prints these preliminary results on a sheet with one line for each day (Figure 2). At the same time a summary of the daily data is stored on magnetic tape. The updating process at the end of the year allows insertion of data for periods not available on the original record, substitution of gage height or discharge figures for periods when unusual hydraulic conditions prevailed, and recomputation of figures of discharge on the basis of more up-to-date information on the stage-discharge relation. The final printout sheet of daily discharges with monthly and yearly summaries is in a form usable as offset manuscript for publication (Figure 3).

ALTERNATE METHODS OF COMPUTING DISCHARGE

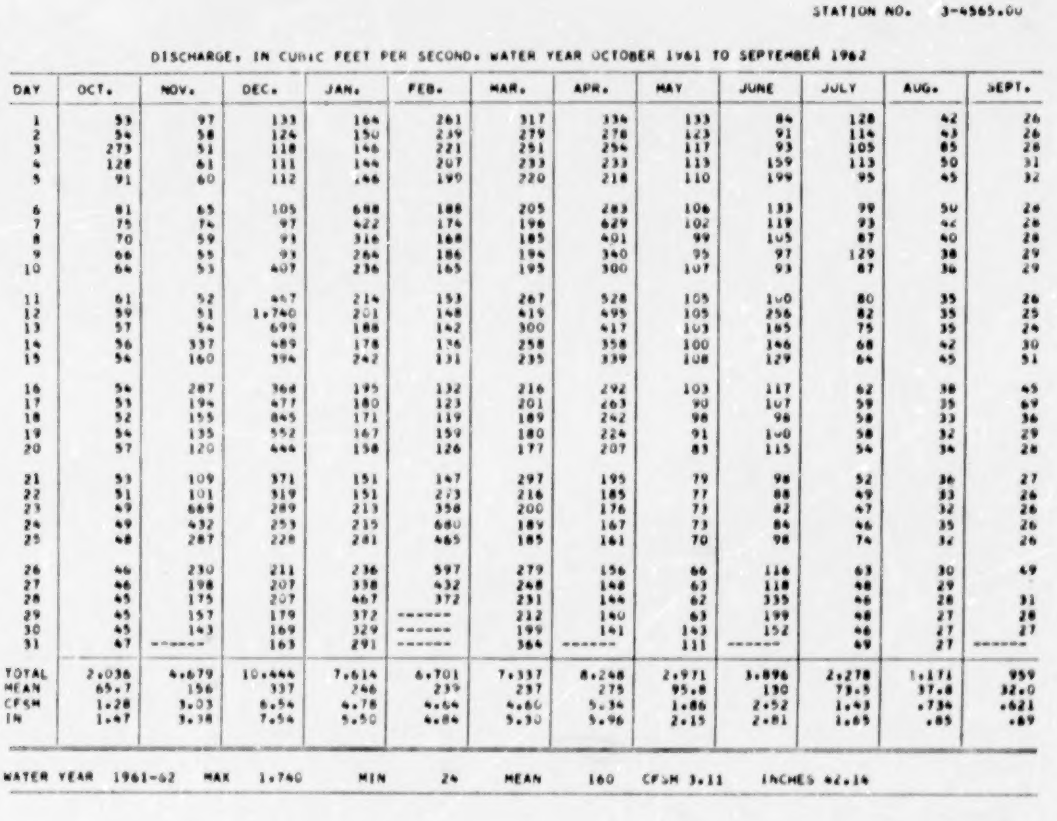
Of course, discharge at all gaging stations cannot be computed by exactly the same method. We started out by programming the computations for the most frequent situation where simple stage-discharge relations can be developed. This type of computation can be used for at least three-quarters of all our gaging stations. Later we added alternative programs for some of the more difficult hydraulic conditions, and we will continue to add other alternative computation methods for other situations from time to time. One of the alternative computations now available is for the so-called "slope stations" where simple stage-discharge relations do not apply. For this type of station, gage heights are recorded by two separate instruments at both ends of a suitable reach of channel, and interrelationships between stage, fall, and discharge are used to compute figures of discharge. This method works well where steady flow or nearly steady flow conditions prevail a major portion of the time. For sites where unsteady flow conditions prevail generally, such as in reaches affected by tide, a much more complicated computa-



```
7-1825.10   NEOSHO RIVER AT BURLINGTON. KANS.                                                    WATER YEAR ENDING SEPT. 30. 1963

RT 21 TO  9-30                                                    BI-HOURLY GAGE HEIGHTS           (SYMBOL TEST DIFF 0.1)

S  DATE  MAX   MIN   MEAN HR EQ-GH DATUM SHIFT MEAN Q  0200 0400 0600 0800 1000 1200 1400 1600 1800 2000 2200 2400 TMAX

A  2-25  6.84  6.50  6.81     6.82               314                                0682 0682 0681 0681 0681 0681 1345
   2-26  6.81  6.81  6.81     6.82               312   0681 0681 0681 0681 0681 0681 0681 0681 0681 0681 0681 0681 2400
   2-27  6.81  6.80  6.80     6.81               309   0681 0681 0681 0681 0681 0680 0680 0680 0680 0680 0680 0680 1000
   2-28  6.80  6.79  6.79     6.80               306   0680 0680 0680 0680 0680 0679 0679 0679 0679 0679 0679 0679 1015

   3-01  6.79  6.78  6.78     6.79               301   0679 0679 0678 0678 0678 0678 0678 0678 0678 0678 0678 0678 0530
   3-02  6.78  6.78  6.78     6.78               300   0678 0678 0678 0678 0678 0678 0678 0678 0678 0678 0678 0678 2400
   3-03  6.79  6.77  6.78     6.78               300   0678 0678 0678 0678 0678 0678 0677 0677 0678 0678 0678 0678 1830
A  3-04  9.60  6.78  7.92     8.07              1020   0678 0678 0680 0709 0732 0754 0801 0872 0908 0921 0932 0960 2400
   3-05 10.33  8.92  9.69     9.69              2190   0993 1018 1032 1033 1019 0996 0967 0938 0915 0898 0892 0900 0800
   3-06 11.61  9.04 10.81    10.83              2850   0926 0966 1011 1056 1097 1126 1151 1160 1159 1153 1135 1114 1730
   3-07 11.11  9.96 10.30    10.30              2540   1089 1067 1048 1031 1022 1015 1013 1009 1008 1004 1000 0996 0015
   3-08 10.68  9.16  9.66     9.68              2180   0988 0980 0969 0958 0944 0932 0919 0916 0947 0987 1025 1068 2400
   3-09 10.94  8.87  9.97     9.99              2360   1094 1087 1064 1040 1019 0998 0978 0958 0936 0918 0900 0887 0315
   3-10  8.85  8.07  8.34     8.35              1240   0873 0863 0851 0843 0834 0829 0823 0819 0816 0811 0808 0809 0015
   3-11  9.02  8.09  8.38     8.40              1280   0812 0814 0816 0819 0821 0825 0831 0842 0854 0870 0886 0902 2400
   3-12 10.22  9.02  9.43     9.45              2050   0911 0914 0917 0918 0921 0928 0939 0951 0966 0982 1002 1022 2400
   3-13 10.95 10.25 10.70    10.71              2780   1045 1066 1080 1088 1094 1095 1090 1082 1072 1057 1044 1028 1245
   3-14 10.27  8.84  9.43     9.45              2050   1013 0994 0979 0964 0950 0936 0925 0916 0905 0899 0891 0884 0015
   3-15  8.84  8.34  8.56     8.57              1430   0879 0873 0868 0863 0860 0854 0852 0847 0844 0840 0835 0834 0015
   3-16  8.34  8.09  8.20     8.22              1130   0833 0829 0827 0824 0821 0821 0818 0816 0815 0813 0811 0809 0045
   3-17  8.09  7.96  8.01     8.02               979   0808 0806 0805 0802 0802 0802 0799 0799 0797 0797 0796 0796 0100
   3-18  7.96  7.85  7.89     7.90               894   0794 0793 0793 0789 0788 0788 0788 0788 0786 0786 0786 0785 0115
   3-19  7.84  7.74  7.79     7.80               825   0784 0783 0783 0781 0780 0778 0778 0776 0776 0774 0774 0774 0300
   3-20  7.72  7.67  7.70     7.71               766   0772 0772 0771 0771 0771 0770 0770 0770 0769 0767 0767 0767 0545
   3-21  7.67  7.64  7.65     7.66               736   0767 0767 0767 0767 0766 0764 0764 0764 0764 0764 0764 0764 0930
   3-22  7.64  7.56  7.61     7.61               706   0763 0763 0763 0763 0760 0760 0760 0760 0759 0758 0758 0756 0115
   3-23  7.56  7.48  7.52     7.52               654   0755 0755 0754 0753 0751 0751 0751 0751 0749 0749 0749 0748 0115
   3-24  7.48  7.44  7.46     7.46               619   0748 0748 0747 0746 0746 0746 0746 0744 0744 0744 0744 0744 0445
   3-25  7.44  7.43  7.43 13  7.44               606   0743 0743 0743 0743 0744 0744                               1300

PERIOD  11.61  6.50
```

Figure 2 — Preliminary Printout Showing Bi-Hourly and Daily Gage Heights and Discharges.

tion has been programmed to obtain figures of discharge. Again, two records of stage are needed, one at each end of the reach. But for each 15-minute time interval, an analysis of the unsteady flow condition is made. This analysis involves an approximate numerical solution of two first-order quasi-linear hyperbolic partial differential equations of two dependent and two independent variables. This type of computation is roughly 10 times as expensive as that for the regular gaging station, but it is presently the only successful method we have for computing flow in tidal reaches of large estuaries. Another program has been developed for "deflection-meter" stations. At these stations a movable vane mounted in a fixed position in a channel gives an index of velocity and direction of flow. Two recorders at the same site are used, one for the stage record and the other for the deflection-meter record. For each 15-minute time period, discharge is computed as the product of an area and a velocity. Since direction of flow is taken into consideration, this method is usable for canals or small streams in tidal reaches as well as for small channels whose very low velocities make it impossible to establish a definite stage-discharge relation and for stations on small channels where conditions other than tides cause changes in flow direction.



```
                                                                                   STATION NO.   3-4565.00

                 DISCHARGE, IN CUBIC FEET PER SECOND, WATER YEAR OCTOBER 1961 TO SEPTEMBER 1962

DAY  | OCT.   | NOV.   | DEC.   | JAN.   | FEB.   | MAR.   | APR.   | MAY    | JUNE   | JULY   | AUG.   | SEPT.
  1  |     53 |     97 |    133 |    164 |    261 |    317 |    336 |    133 |     84 |    128 |     42 |     26
  2  |     54 |     58 |    124 |    150 |    239 |    279 |    278 |    123 |     91 |    114 |     43 |     26
  3  |    273 |     51 |    118 |    146 |    221 |    251 |    254 |    117 |     93 |    105 |     85 |     28
  4  |    128 |     61 |    111 |    144 |    207 |    233 |    233 |    113 |    159 |    113 |     50 |     31
  5  |     91 |     60 |    112 |    146 |    199 |    220 |    218 |    110 |    199 |     95 |     45 |     32

  6  |     81 |     65 |    105 |    688 |    188 |    205 |    283 |    106 |    133 |     99 |     50 |     28
  7  |     75 |     74 |     97 |    422 |    174 |    196 |    629 |    102 |    119 |     93 |     42 |     28
  8  |     70 |     59 |     93 |    316 |    168 |    185 |    401 |     99 |    105 |     87 |     40 |     28
  9  |     66 |     55 |     93 |    264 |    186 |    194 |    340 |     95 |     97 |    129 |     38 |     29
 10  |     64 |     53 |    407 |    236 |    165 |    195 |    300 |    107 |     93 |     87 |     36 |     29

 11  |     61 |     52 |    447 |    214 |    153 |    267 |    528 |    105 |    140 |     80 |     35 |     26
 12  |     59 |     51 |  1,740 |    201 |    148 |    419 |    495 |    105 |    256 |     82 |     35 |     25
 13  |     57 |     54 |    699 |    188 |    142 |    300 |    417 |    103 |    185 |     75 |     35 |     24
 14  |     56 |    337 |    489 |    178 |    136 |    258 |    358 |    100 |    146 |     68 |     42 |     30
 15  |     54 |    160 |    394 |    242 |    131 |    235 |    339 |    108 |    129 |     64 |     45 |     51

 16  |     54 |    287 |    368 |    195 |    132 |    216 |    292 |    103 |    117 |     62 |     38 |     45
 17  |     53 |    194 |    477 |    180 |    123 |    201 |    263 |     90 |    107 |     59 |     35 |     69
 18  |     52 |    155 |    845 |    171 |    119 |    189 |    242 |     98 |     96 |     58 |     33 |     36
 19  |     54 |    135 |    552 |    167 |    159 |    180 |    224 |     91 |    140 |     58 |     32 |     29
 20  |     57 |    120 |    444 |    158 |    126 |    177 |    207 |     83 |    115 |     54 |     34 |     28

 21  |     53 |    109 |    371 |    151 |    147 |    297 |    195 |     79 |     98 |     52 |     36 |     27
 22  |     51 |    101 |    319 |    151 |    273 |    216 |    185 |     77 |     88 |     49 |     33 |     26
 23  |     49 |    469 |    289 |    213 |    358 |    200 |    176 |     73 |     82 |     47 |     32 |     26
 24  |     49 |    432 |    253 |    215 |    680 |    189 |    167 |     73 |     84 |     46 |     35 |     26
 25  |     48 |    287 |    228 |    281 |    465 |    185 |    161 |     70 |     98 |     54 |     32 |     26

 26  |     46 |    230 |    211 |    236 |    597 |    279 |    156 |     66 |    116 |     63 |     30 |     49
 27  |     46 |    198 |    207 |    338 |    432 |    248 |    148 |     63 |    118 |     48 |     29 |
 28  |     45 |    175 |    207 |    467 |    372 |    231 |    144 |     62 |    335 |     46 |     28 |     31
 29  |     45 |    157 |    179 |    372 | ------ |    212 |    140 |     63 |    199 |     48 |     27 |     28
 30  |     45 |    143 |    169 |    329 | ------ |    199 |    141 |    143 |    152 |     46 |     27 |     27
 31  |     47 | ------ |    163 |    291 | ------ |    364 | ------ |    111 | ------ |     69 |     27 | ------

TOTAL|  2,036 |  4,679 | 10,444 |  7,614 |  6,701 |  7,337 |  8,248 |  2,971 |  3,896 |  2,278 |  1,171 |    959
MEAN |   65.7 |    156 |    337 |    246 |    239 |    237 |    275 |   95.8 |    130 |   73.5 |   37.8 |   32.0
CFSM |   1.28 |   3.03 |   6.54 |   4.78 |   4.64 |   4.60 |   5.34 |   1.86 |   2.52 |   1.43 |   .734 |   .621
IN   |   1.47 |   3.38 |   7.54 |   5.50 |   4.84 |   5.30 |   5.96 |   2.15 |   2.81 |   1.65 |    .85 |    .69

WATER YEAR  1961-62   MAX   1,740      MIN      24     MEAN     160   CFSM 3.11    INCHES 42.14
```

Figure 3 — Final Printout Showing Daily Discharges with Monthly and Yearly Summaries.

MODIFICATIONS OF THE STANDARD DIGITAL RECORDER

In all the applications mentioned so far, the standard digital recorder with mechanical input is used without modifications. A few of these standard recorders have been modified slightly by the Geological Survey for use as precipitation gages. In this simple digital precipitation gage, a small float attached to a pulley on the digital recorder input shaft measures accumulated depth of water in the rain gage reservoir. Incidentally, the Weather Bureau has recently sponsored development of a much more elaborate weighing rain gage in which the same basic recording techniques and the same type of punched output tape are used.

The standard recorder with very minor modification is being used for measuring depths to ground water at about 40 points in a project in Arizona where a detailed coordination with simultaneous records of soil moisture and nearby stream flow records is required.

The only major modification of the standard recorder made by the Geological Survey has been the adaptation of the basic recorder to accommodate electrical rather than mechanical inputs. This is important because sensing devices for chemical quality items generally have electrical outputs. Also, with electrical input, it only takes a small additional modification to allow for multiple inputs. The modification for electrical input involves using a positioning motor on the input shaft in place of the direct mechanical input. If multiple inputs are desired, a stepping switch can be inserted to connect the recorder to each electrical input sensor in turn at the time that each set of readings is desired. The Geological Survey has added such modifications to a few standard recorders. For instance, one recorder has been adapted by our Quality of Water office in Florida to record four items consisting of top and bottom temperature and conductivity in a particular stream. The manufacturer can supply units already adapted for single or multiple electrical inputs.

In addition to the modifications for different kinds of input, a telemetering device can be attached to the standard recorder. In fact, the basic recorder was specifically designed with this option in mind. Space was provided just behind the punch block for a set of contacts that are operated by the punch pins at the moment of punching. A wiring cable brings this information to an external box containing circuitry to store the last punched information until the next reading and to decode the information from binary-decimal to straight-decimal and to transmit it as a sequence of recognizable tones to telephone or radio transmitting equipment. A number of the telemetering attachments with telephone transmitting equipment are being added to our river gages in a cooperative project with the Weather Bureau. Certain of our river gages now equipped with digital recorders can be used by the Weather Bureau for flood forecasting.

CONCLUSION

No doubt, there are other, as yet unconsidered, possibilities for use of this same basic field recorder in hydrologic investigations. But our uses of this recorder and the computer processing techniques for streamflow data alone save us significant amounts of manpower at regular stations and enable us to obtain flow data in places where we could not obtain the information in any other manner. The release of some of our technical manpower from the drudgery of routine data processing allows us to do more interpretive work that will lead more directly to the solution of specific water problems and to the expansion of our general knowledge of the behavior of water in nature.

DISCUSSION

Mr. Isherwood indicated that three translator systems are being used in the system. The first two cost about $6,000 each and the third, which operates four times faster, costs about $12,000. It is expected that a magnetic tape translator that operates 10 times faster than the last will be needed soon and will cost about $15,000.

He indicated that the group has used three different computers. In each case they started out at a low use rate but within 3 to 4 years were operating 24 hours per day; they expect to need a larger unit in the near future. The paper tapes are being read by a standard photoreader that reads about 1000 characters per second; it rents for about $14,000 per month. When thousands of stations are in operation, however, this reader will be too slow. They plan then to use magnetic tape that can be read at 25,000 to 50,000 characters per second.

In response to a question about the use of paper tape under high humidity, Mr. Isherwood indicated that the papers have held up remarkably well. The National Bureau of Standards indicated that so-called waterproof papers get just as wet as standard papers, but not as quickly.

Recently, a foil-backed paper with more mechanical strength has been introduced. Mr. Isherwood further indicated that the electrical components are more sensitive to moisture than the mechanical parts of the system.

John J. Gannon
Associate Professor of Public Health Engineering
School of Public Health
University of Michigan, Ann Arbor

SUMMARY

Several graphical and statistical procedures are presented for the interpretation and analysis of hydrologic data from the standpoint of influence on water quality. These include the hydrograph; development of seasonal patterns on normal and log-normal probability papers; analysis of drought flows with examples of procedures used in a Michigan study; determination of time of passage by displacement calculations and tracer methodology; and comparison of potential regulated flows and natural flows. These procedures used with good judgment have proved their usefulness in many water pollution investigations.

THE INTERPRETATION AND ANALYSIS OF HYDROLOGIC DATA

Several graphical and statistical procedures are available for the interpretation and analysis of hydrologic data from the standpoint of influence on water quality. It is the intent of this paper to present and discuss the use of the hydrograph — both continuous and daily average; normal and log-normal probability papers and their use in developing seasonal patterns; log-extremal probability paper and its use in analysis of drought flows, including certain adjusting and summary procedures; systematic studies of the influence of flow regulation; and the importance of knowledge of the physical characteristics of the river channel and the use of this information in determination of river time of passage. Certainly, this is not an exhaustive list of all hydrologic considerations, but these graphical and statistical procedures have proved useful in many water pollution investigations.

THE HYDROGRAPH

The hydrograph consists of a graph of time versus river flow at a particular location. The time scale in some cases is presented on a continuous basis, resulting in an instantaneous hydrograph frequently expressed in terms of river stage rather than runoff. In other cases, the time scale is presented on a daily basis, with the flow averaged over the day resulting in a daily hydrograph. In still others, it is presented on a monthly basis, with the flow averaged over the month resulting in a monthly hydrograph. Each has its advantages and disadvantages. The continuous and daily-average hydrographs are presented and discussed below; the monthly average hydrograph is considered later in connection with the development of seasonal patterns of runoff.

THE CONTINUOUS HYDROGRAPH

Most of the important stream-gaging stations maintained by the U. S. Geological Survey in the United States operate on a continuous basis, generally making a continuous recording of river stage with time. This river stage must, of course, be converted to discharge, by means of an appropriate rating curve. In river situations where there are diurnal fluctuations in flow — frequently induced by activities of man — it is important to know of these fluctuations and to minimize their effect in the design of any stream sampling program. Here is where a continuous-gage chart can be extremely valuable.

Unfortunately, these charts are not routinely published and must be obtained from the files of the appropriate district engineer's office of the Geological Survey.

In two recent intensive stream surveys conducted by the writer, hour-to-hour fluctuations in stream flow through the critical reach of river were observed during the dry warm-weather period of August. This is the ideal period for evaluation of a water quality problem related to organic wastes.

The first survey involved the Clinton River in Michigan and covered the section of the river from below the Pontiac waste treatment plant outfall to the village of Rochester, a distance of 11.41 river miles. Fortunately, a Geological Survey continuous-recording stream-gaging station is located at Auburn Heights in the critical stretch 3.02 river miles below the Pontiac waste treatment plant outfall. Figure 1 shows the Auburn Heights continuous-gage chart for the period August 17 through 25, 1960. It can be seen that a definite hour-to-hour flow fluctuation exists and is produced primarily by flow variation from the Pontiac waste treatment plant. Average flow during the intensive 48-hour sampling period, August 23, 24, and 25, 1960, was 33 cfs at Auburn Heights, while the average flow from the waste treatment plant was 16.9 cfs; thus, the treatment plant effluent made up more than 50 percent of the total river flow during this period. Also, it can be seen that while the average flow was 33 cfs, the actual flow ranged from 26 to 46 cfs. Although it was not possible to alter the flow pattern it was possible to minimize this influence by collecting river samples every 4 hours around the clock for a 48-hour period through the critical section.

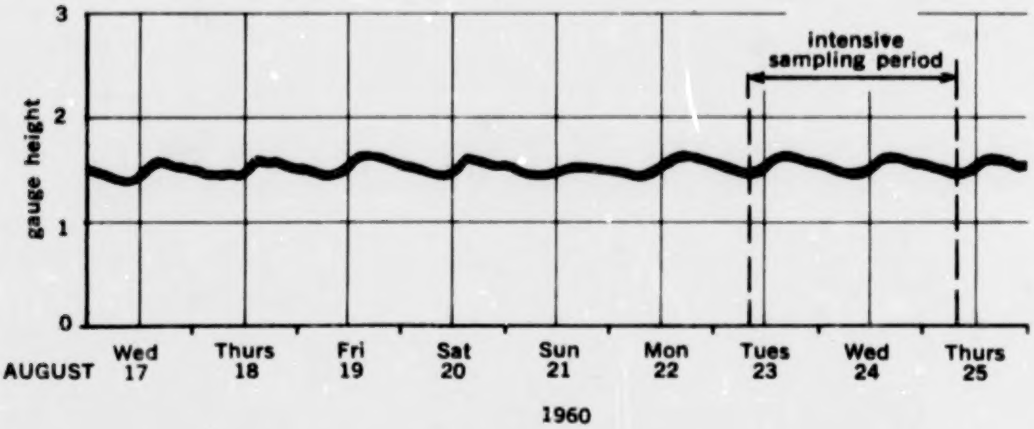


Figure 1 — Gage Chart for Clinton River at Auburn Heights, August 17 through 25, 1960.

The second survey involved the Tittabawassee River in Michigan and included the section from below Midland and the Dow Chemical Company waste treatment plant outfall to Saginaw, a distance of 19.25 river miles. Fortunately, as in the previously mentioned cases, a Geological Survey continuous-recording gage is located on the Tittabawassee River at Midland opposite the grounds of the Dow Chemical Company. Figure 2 shows the Midland continuous-gage chart for the period August 17 through 25, 1961. Diurnal fluctuations in flow are induced during weekdays by a hydropower installation upstream from Midland, and these fluctuations are illustrated by the usual variations on August 17 and 18. For purposes of conducting an intensive stream sampling program under steady-flow conditions below Midland, arrangements were made with the hydropower company to lower their reservoir on August 21 and 22, and then to stop operations and hold back the river flow for 48 hours August 23, 24, and 25.

1961, thus, creating an artificial drought condition. This proved to be an extremely successful operation, resulting in the accumulation of a considerable amount of useful data in a short time, under favorable flow conditions.

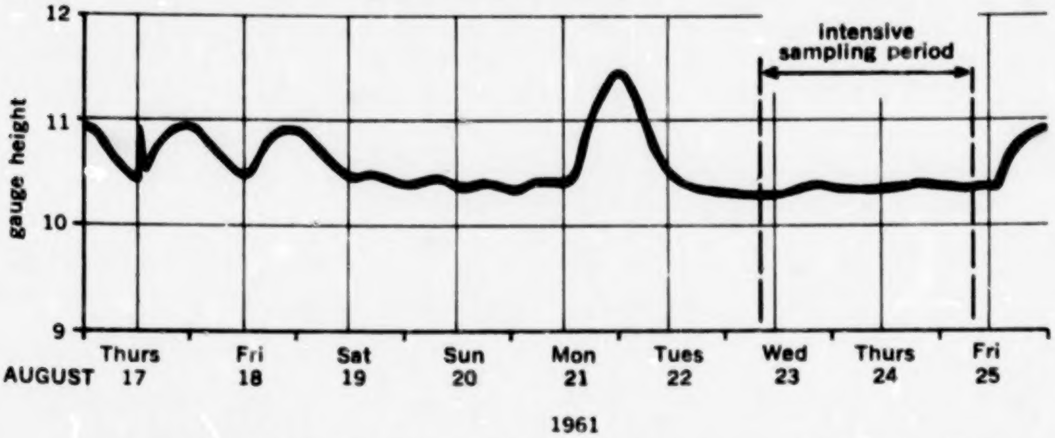


Figure 2 — Gage Chart for Tittabawassee River at Midland, August 17 through 25, 1961.

It should be reported that a certain amount of the Tittabawassee River water is diverted around the Midland gage by the Dow Chemical Company, and although this adds to the total river flow below Dow, it is of a steady nature and does not induce further pulsations in flow. The diversion will be discussed later.

THE DAILY HYDROGRAPH

For many purposes a daily hydrograph of the daily average flow versus time is useful. Generally, the daily average flow is the shortest period of flow regularly reported in the Water Supply Papers[1] of the Geological Survey and is, therefore, readily available for all published stations. The daily hydrograph is useful in characterizing a river as "flashy" or "stable," i.e., rapid change in flow from day to day or, gradual change from day to day. Also, it can be useful in relating water quality data to the flow conditions that prevailed during the sampling period, including such things as high or low runoff, rising or falling hydrograph, and stable or unstable flow conditions.

Figure 3 is the daily hydrograph for 1961 for the Clinton River at Auburn Heights, Michigan. It might be characterized as a "flashy" hydrograph resulting from several drainage area characteristics, including its small size of 123 square miles. Routine sampling days are indicated across the top of the hydrograph, allowing an immediate visual comparison of runoff conditions during and preceding these sampling periods.

In contrast, Figure 4 is the daily hydrograph for 1951 for the Savannah River near Clyo, Georgia, which might be characterized as a "stable" hydrograph. Probably, one of the most important factors contributing to this stability is the large drainage area of 9850 square miles. A period of intensive water quality sampling is indicated in August 1951 when the runoff during and preceding the sampling period was relatively stable.

Just as the daily hydrograph can be useful in relating previous runoff and water quality conditions, it can also be helpful in planning stream surveys, especially if they are to be the intensive type conducted over short periods of time under steady, low runoff conditions. Plotting and studying daily hydrographs for a particular river location for several years preceding a planned stream survey period tend to identify the time of the year most likely to have a steady, low-flow condition. This then serves as a guide for the assembly of the necessary sampling personnel and equipment, together with the supporting laboratory facilities. As the planned survey period approaches, the maintenance of a current daily hydrograph, together with knowledge of the weather forecasts for the survey period, enables the investigator to know whether runoff conditions are approaching an acceptable level, and also, whether there is a reasonable chance of a dry period that will result in a steady flow condition. This approach has been useful in planning several stream surveys, and generally, results in the accumulation of considerable data under desirable runoff conditions.

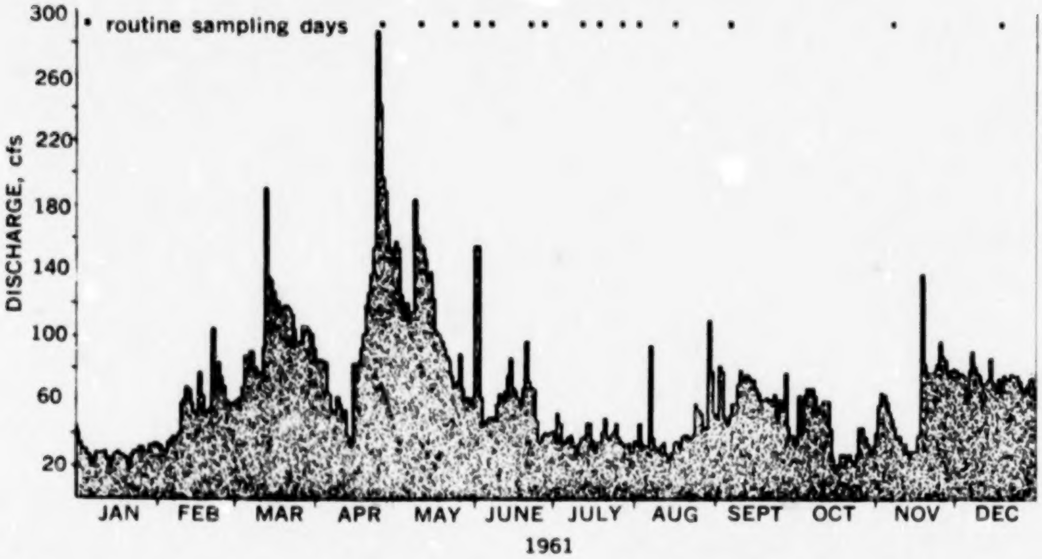


Figure 3 — Daily Hydrograph for Clinton River at Auburn Heights, Michigan, a 123-square-mile Drainage Area.

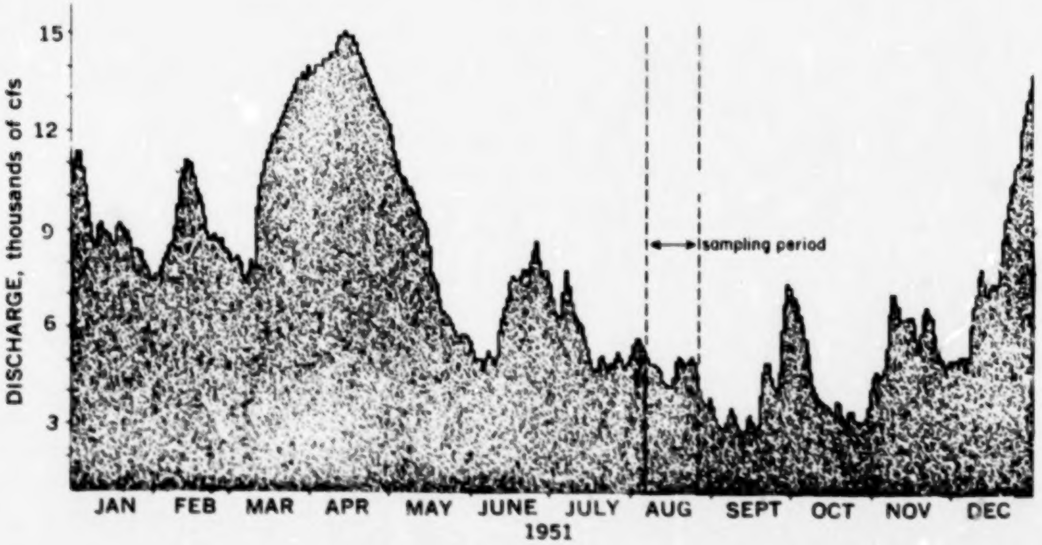


Figure 4 — Daily Hydrograph for Savannah River near Clyo, Georgia, a 9850-square-mile Drainage Area.

PROBABILITY PAPERS AND SEASONAL PATTERNS

Probability paper facilitates the application of statistical theory in summarizing data, and several types including normal, logarithmic normal, linear extremal, and logarithmic extremal, have been used in summarizing hydrological and meteorological observations. The early work of Hazen[2] and the more recent work of Velz[3] are to be particularly noted. This section of the paper will deal with normal and log-normal probability papers, and their application in the development of seasonal patterns of selected hydrologic phenomena.

NORMAL PROBABILITY PAPER

Graphical methods describing the relationships expressed by the normal distribution are available in terms of normal probability paper. Here is a quick and easy procedure that makes available most of the advantages of the statistical method.

Briefly, normal probability paper is constructed by summing the area under the normal probability curve from left to right, thereby obtaining an expression for the x or horizontal axis of percent equal to or less than. Such a grid is illustrated in Figure 5, where it is noted that a clustering of percentage occurs around the 50 percent or centering value, with a considerable spread toward the upper and lower end of the scale. Further, a definite relation is observed between the standard deviation (σ) scale across the top and the percent equal to or less than scale across the bottom following the normal distribution. The vertical, or y scale, is linear and is assigned the units of measurement of the observations involved.

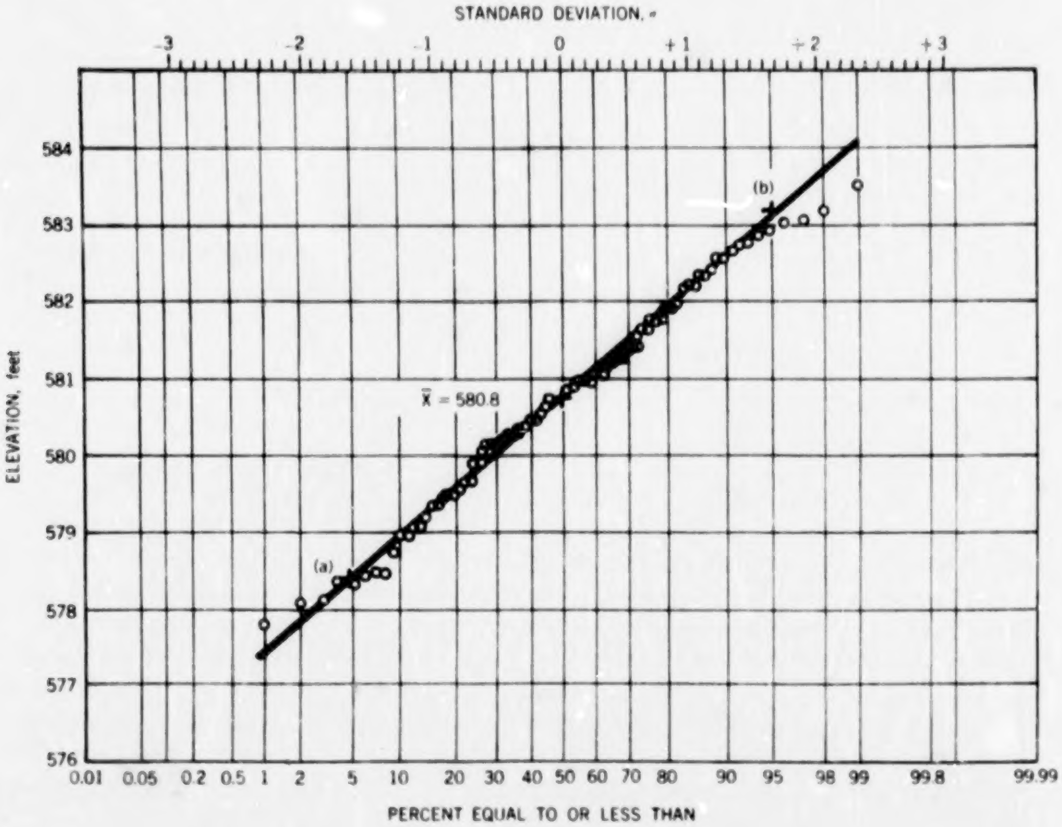


Figure 5 — Monthly Mean Elevation for the Lake-Michigan-Huron System for May 1860 through 1957.

Data that follow the normal probability curve plot as a straight line on this grid; thus, there is available a quick method of testing the normality of a series of observations. Furthermore, the slope of the line is a measure of variation, the steeper the slope the more variation, the flatter the slope the less variation. There is a definite relationship between the slope of the line and the standard deviation, making it possible to determine standard deviation graphically.

A more complete discussion of normal probability paper has been presented elsewhere by Gannon[4] and Velz,[3] including the mechanics of plotting on the grid, and will not be repeated here.

Figure 5 is an illustration of the application of normal probability paper in defining the variation of the monthly mean lake level for May for the Lake Michigan-Huron system for the period of record 1860 through 1957. This illustration is taken from a recent publication of Velz and Gannon.[5]

In Figure 5 it can be seen that the points describe a straight line, thus, indicating that the data are normally distributed. Furthermore, from a statistical standpoint it is possible to graphically determine the mean ($\overline{X}$), which in this case has an elevation of 580.8 feet. In addition to the mean, it is possible to define variation around the mean, such as the 90 percent confidence range, i.e., 90 percent of the individual values fall within this range around the mean, while 5 percent are less than the lower limit, and 5 percent are greater than the higher limit. The lower limit indicated at point (a) of the distribution opposite the 5 percent equal to or less than line is seen to have a value of 578.4 feet, while the upper limit indicated at point (b) of the distribution opposite the 95 percent equal to or less than line is seen to have a value of 583.1 feet. Thus, normal probability paper has been useful in defining the mean monthly lake level during May for the period of record, together with the 90 percent confidence limit of these monthly values.

SEASONAL PATTERN OF LAKE LEVELS

By an analysis similar to that indicated in Figure 5 for each month of the year, it is possible to develop a seasonal pattern of lake levels such as that illustrated in Figure 6 for the Lake Michigan-Huron system. Curve A is seen to be the most probable monthly average lake level, while the 90 percent confidence range around this most probable value is indicated by a dashed line. It is interesting that the low monthly average level usually occurs in the winter months of January and February, whereas the high monthly average level generally occurs in the summer months of July or August. In addition to the 90 percent confidence range, the highest and lowest observed monthly average level for the period of record is indicated for each month. Thus, there is available in a single chart most of the important summary data for the lake levels of the Michigan-Huron system.

LOGARITHMIC NORMAL PROBABILITY PAPER

Certain types of data do not plot as a straight line on normal probability paper; however, in some cases, they straighten out on a logarithmic vertical scale. In some instances, it is not possible to anticipate whether the data will follow a normal or a logarithmic normal distribution, and the only practical solution is to try both. A typical logarithmic normal probability grid is illustrated in Figure 7, where it can be seen that the probability scale is the same as it would be on normal paper, whereas the vertical scale is logarithmic instead of linear. The mechanics of plotting on logarithmic probabilty paper are the same as for normal probability paper, and generally the same type of information is obtained.

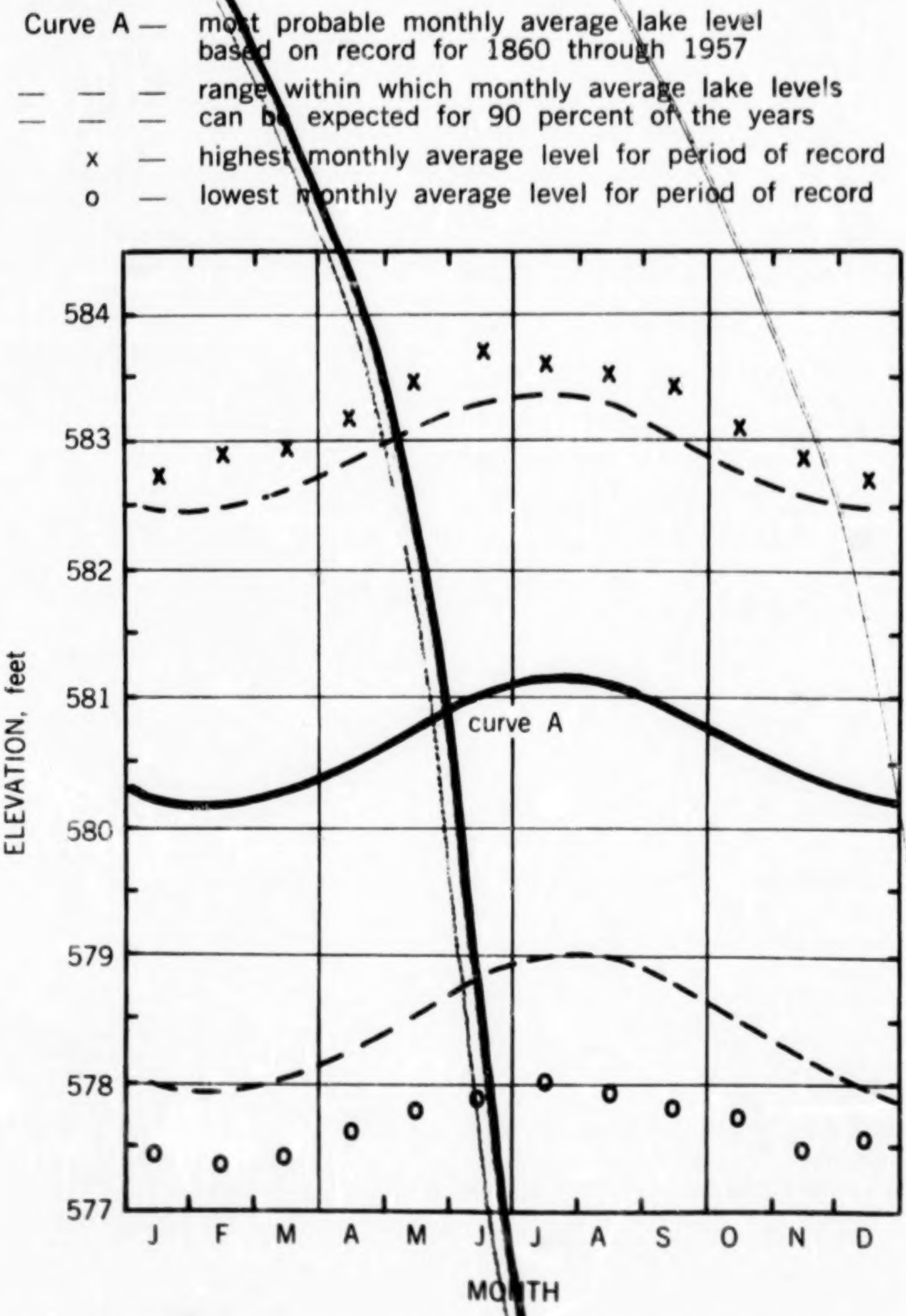


Figure 6 — Seasonal Pattern of Levels for Lake Michigan-Huron System.

Figure 7 is an illustration of the application of logarithmic normal probability paper in defining the variation of the monthly average flow for May for the Kalamazoo River at Comstock, Michigan, for the period of record, October 1935 to September 1960. Generally, experience has indicated that monthly average runoff figures are best described by a logarithmic normal distribution, but there is no fundamental explanation why.

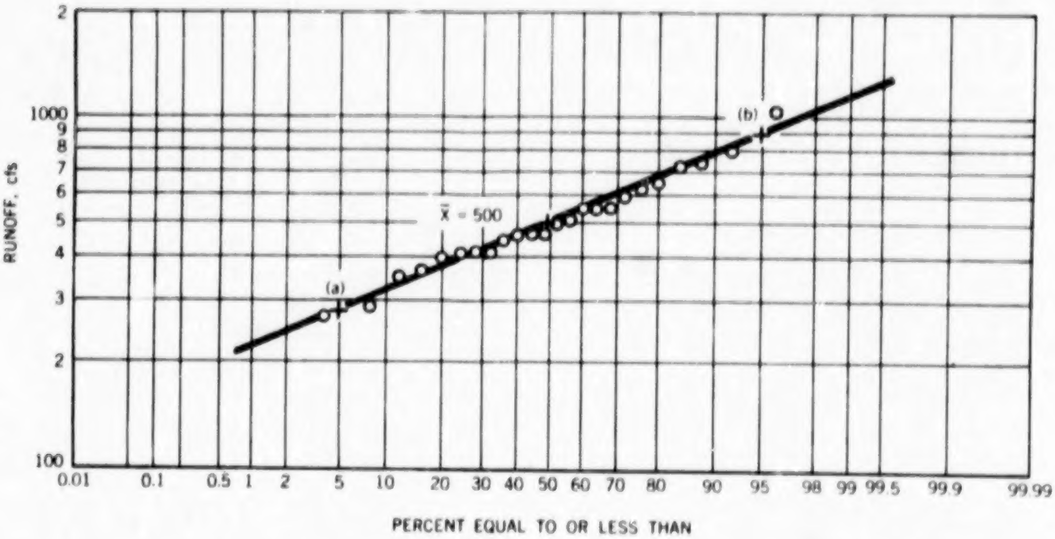


Figure 7 — May Monthly Average Flow of Kalamazoo River at Comstock.

In Figure 7 it can be seen that the points describe a straight line, thus, indicating that the data are logarithmically normally distributed. As with normal probability paper, it is possible to determine graphically the mean ($\bar{X}$), which in this case has a value of 500 cfs, whereas the 90 percent confidence range around the mean is indicated at points (a) and (b), which have values of 280 and 900 cfs, respectively.

SEASONAL PATTERN OF RUNOFF

Just as it is possible to develop a seasonal pattern of lake levels, so is it possible to develop a seasonal pattern of runoff. Figure 8 is an illustration of such a chart for the Kalamazoo River at Comstock, developed from an analysis of the variation in monthly average flow for each month of the year on logarithmic normal probability paper, as illustrated in Figure 7. Such a figure is in effect a type of monthly hydrograph.

It is seen that Curve B in Figure 8 is the most probable monthly average flow and Curve A the mean for the period of record. In addition, the dashed lines C and D indicate the 90 percent confidence limits of individual monthly values around the most probable. For the Kalamazoo River, the high runoff period occurs in the spring months of March and April, whereas the low runoff period occurs in the late summer months of August and September.

In contrast to the Kalamazoo River in Michigan, Figure 9 illustrates the seasonal pattern of runoff for the Platte River at Sinclair, Wyoming, for the period 1940 through 1961. The high flow generally occurs in June whereas the low flow occurs in September and again in January. Also, the variation from month to month is greater than for the Kalamazoo River.

The graph showing seasonal pattern of runoff is helpful in depicting the most probable flow available each month of the year, together with its variation, rather than the lowest flows only. If one subscribes to the concept of using the total river flow for waste assimilation purposes, either by means of storing the high river flow in reservoirs and releasing it during the low flow periods or of storing the waste by means of storage lagoons and releasing this waste in accordance with river flow, then flow information of the type presented in the seasonal pattern of runoff is essential.

DROUGHT FLOW ANALYSIS

To meet the need for knowledge concerning the probability of occurrence of drought flows, particularly as they relate to water quality considerations, special graphical pro-

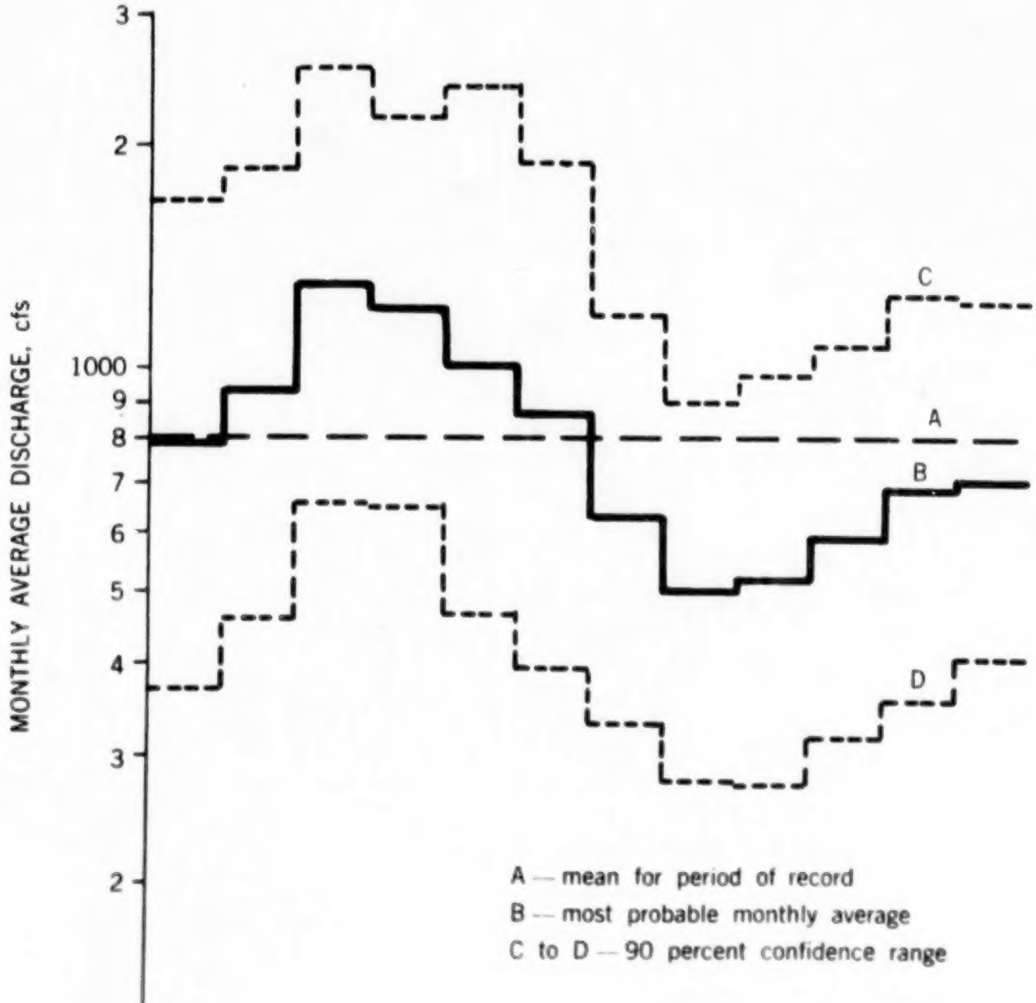


Figure 8 — Seasonal Pattern of Runoff for Kalamazoo River at Comstock, Monthly Average Discharge, October 1935 through September 1960.

cedures have been developed and adapted employing the theory of extreme values as proposed by Gumbel.[6-8] These procedures have been successfully used in a state-wide analysis of the drought flows of the streams of the State of Michigan to compile a comprehensive report on the subject by Velz and Gannon.[5] This section of the paper will discuss the analysis of drought flows on logarithmic extremal probability paper, including certain adjusting and summary procedures together with a consideration of natural and artificial influences. Where it is necessary to establish river water quality standards, these standards should be related to drought flow levels, and information on the probability of these flow levels is essential.

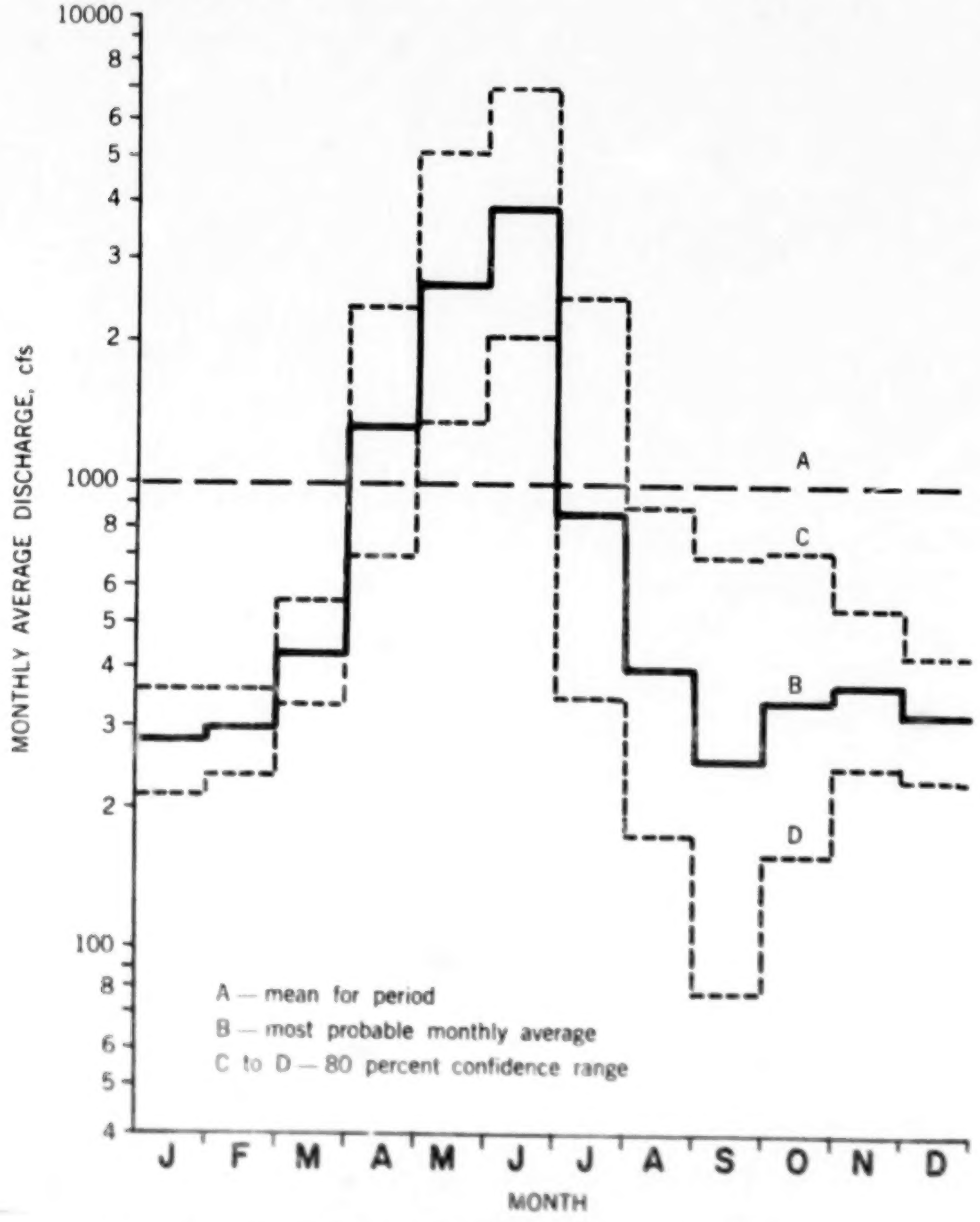


Figure 9 — Seasonal Pattern of Runoff for Platte River at Sinclair, Wyoming. Monthly Average Discharge.

LOGARITHMIC EXTREMAL PROBABILITY PAPER

The extremes of hydrologic observations such as floods and droughts do not follow a normal symmetrical distribution but rather are skewed (the more severe values deviate beyond the mean to a much greater extent than the less severe values deviate below it). Gumbel[6–8] has proposed three asymptotic probabilities of extremes suggesting that the third asymptotic distribution is suitable for analyzing droughts. In the Michigan drought study, logarithmic extremal probability paper was used and the third asymptotic distribution of smallest values followed as suggested by Gumbel.

Such a grid is illustrated in Figure 10, which was developed in a manner similar to those for normal and logarithmic normal probability papers previously discussed. The probability equal to or less than scale is unbalanced to the left, with the more severe values to the right having the greatest spread in accordance with the skewed nature of the distribution. Also, an additional scale has been added across the top, called the return period (T), which is related to the probability scale across the bottom; this scale is particularly useful in dealing with hydrologic data. For example, if the base unit of time from which low flow data are selected is a year, then the return period of 10 would indicate a 1-in-10-year drought.

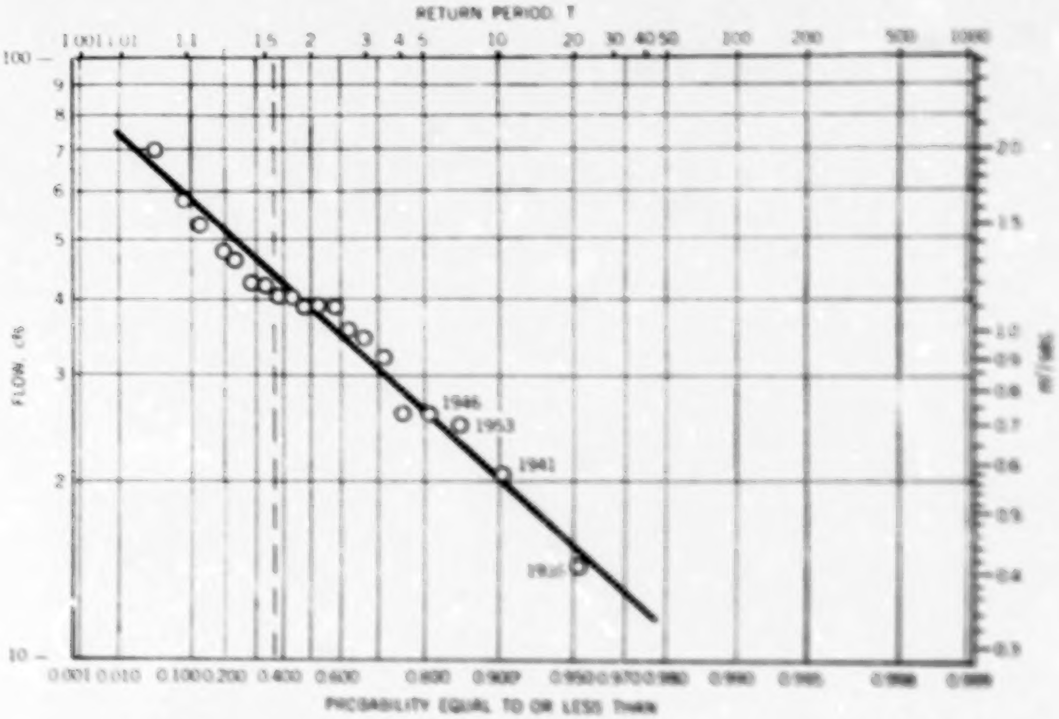


Figure 10 — Minimum 30-day Flow during May through October in Grand River at Jackson, Michigan (Gumbel's Logarithmic Extremal Probability Paper).

The vertical, or y scale, is logarithmic and is assigned the units of measurement involved such as cubic feet per second (cfs). For purposes of use, data are arranged and plotted in order of severity, which in the case of low flows means ordering from the higher to the lower absolute values. In addition, it is necessary to calculate a plotting position for the probability scale using Gumbel's refinement as illustrated by Velz,[3] or Velz and Gannon.[9]

In Figure 10 it is seen that the minimum consecutive 30-day flows for the Grand River at Jackson, Michigan, do approximate a straight line, with the straight line fitted to the data by eye rather than by use of a more rigid mathematical method. Since a good straight line fit results, there is an indication that these data do follow the third asymptotic distribution of smallest values. From the fitted line and the return period scale, the most probable minimum consecutive 30-day drought is read as 44 cfs (more properly called the characteristic drought indicated as a dashed vertical line), the 1-in-5-year drought as 26 cfs, the 1-in-10-year drought as 20.5 cfs, and the 1-in-20-year drought as 16 cfs.

Not all drought flow data necessarily plot as a straight line on logarithmic extremal probability paper, especially if storage, either artificial or natural, is involved or if flow augmentation is involved. An adjusting procedure to handle these cases will be discussed subsequently in this paper.

BASIC INFORMATION

Three time elements are involved in the definition of drought flow: (1) the base unit of time from which a low flow is selected from the record, (2) the length of time over which a low flow is averaged, and (3) the season in which the selection is made.

Ideally, from a statistical standpoint, extreme values selected from consecutive time units should be independent of each other. With low flows, there is a possibility of a carry-over influence from one year to the next, but notwithstanding this possible influence, the base unit of time of the year was used in the Michigan study, primarily because of the relatively short records available in this State.

The second time element, the length of time over which a flow is averaged, may vary depending on the particular application intended for the information. To meet as many time needs as possible, the Michigan study has reported and analyzed five flow periods: the minimum day; the minimum consecutive 7-day, 15-day, and 30-day averages; and the minimum calendar monthly average.

The third time period, the season, is important in differentiating between warm-weather and cold-weather droughts. Generally speaking, in Michigan, cold-weather droughts are different from warm-weather droughts. Because main interest in this study was in connection with warm-weather applications such as water pollution control, irrigation, and recreation uses, low flows were selected from the summer-fall period May 1 through October 31. The base unit of time of the year was retained, but low-flow selections were made only from this summer period.

DROUGHT DURATION VERSUS SEVERITY

As a summary device and as an interpolating aid, a chart similar to Figure 11 has been prepared for each gage to show the relationship between drought duration and severity. Information for the construction of this chart was obtained from four separate logarithmic extremal probability plots similar to Figure 10, covering the minimum daily flows and the minimum consecutive 7-day, 15-day, and 30-day averages. From each plot, the most probable, the 1-in-5-year, the 1-in-10-year, and the 1-in-20-year figures were obtained; these served as a basis for the development of the most probable, the 1-in-5-year, 1-in-10-year, and 1-in-20-year curves in Figure 11. These curves then serve as a framework from which a drought of any duration from 1 to 30 consecutive days can be determined for the indicated return periods. Because of the influence of regulation, the minimum daily flows are in many instances out of line with the rest of the data. To caution the user of this fact and to urge care in the use of data in this short duration range for interpolation purposes, the 1-day and 7-day duration points have been connected by a dashed line.

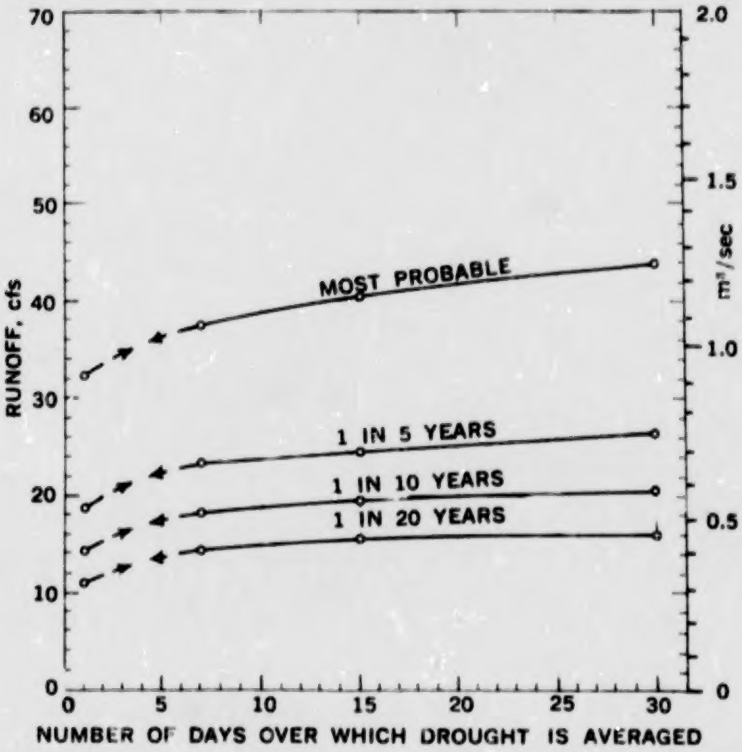


Figure 11 — Chart of Drought Duration Versus Severity for Grand River at Jackson, Michigan.

ADJUSTING PROCEDURES

Not all of the logarithmic extremal probability plots developed as straight lines, especially where regulation was involved, either artificial or natural. In several instances, because of the presence of a base flow below which the river flow had not fallen, a curve developed when the original data were plotted on probability paper. Such a case is illustrated by the solid points in Figure 12, for the minimum daily flows for the Kalamazoo River near Battle Creek. Gumbel[6] has proposed an elaborate computational procedure, involving the third moment of the distribution, for the evaluation of the lower limit and for fitting a curve through the data.

A much simpler technique involves estimation of the base flow by eye, subtracting this figure from each flow, and replotting the remainders as illustrated in Figure 12. If a straight line does not develop, a second and a third estimate of the base flow is made until a straight line results. Thus, in a relatively few trials it is possible to estimate the base flow, and also, to fit a straight line to the remainder. From this line, it is possible to determine the most probable, the 1-in-5-year, 1-in-10-year, and 1-in-20-year flows to which must be added the previously subtracted base flows to bring the flow figures back to their original levels.

For the illustration of Figure 12, it is seen that the base flow was estimated as 135 cfs and that when this flow was subtracted from the original flows the remainders formed a reasonably straight line.

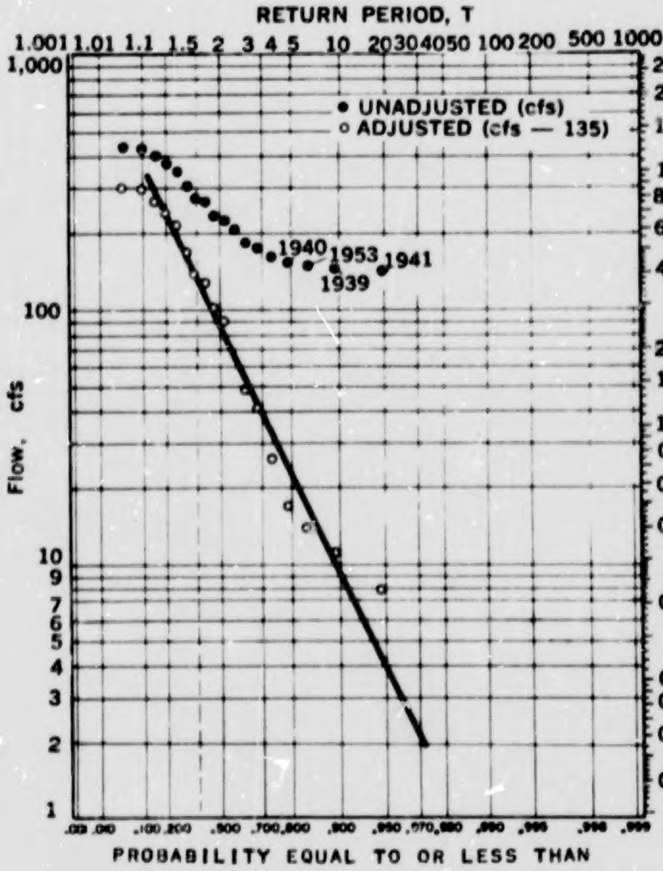


Figure 12 — Minimum Daily Flow during May through October in Kalamazoo River near Battle Creek (Gumbel's Logarithmic Extremal Probability Paper).

DROUGHT FLOW INDICES

Two important drought flow indices have been developed for all of the gages studied in Michigan, namely, the yield and the variability ratio. These summary figures allow comparison of gages within a basin, and in addition, allow comparison of the flow characteristics of one basin with another.

The yield, defined as the discharge per unit drainage area (cfs/mi^2), is useful in reducing discharge figures at gages with varying drainage area sizes to a common base. Considerable variation was observed in the yield characteristics of the several basins. For example, the Manistee River near Sherman, with a drainage area of 900 square miles, shows a 1-in-10-year drought as a 7-day average of very high yield, about 0.8 cfs per square mile. In contrast, the Raisin River at Monroe, with a drainage area of 1034 square miles, shows a very poor yield, about 0.03 cfs per square mile.

The secondary summary index, the variability ratio, defined as the ratio of the 1-in-10-year drought to the most probable drought, is helpful in defining the variation that can be expected in drought flows from year to year. Because of the nature of logarithmic extremal probability paper, and also, because of the adjusting procedures used in some cases, the conventional measures of variation, such as the standard deviation, are not applicable, and it became necessary to develop a new measure. To meet this need, Velz and Gannon[5] proposed the variability ratio, which is easy to determine and which serves as a basis for comparison among gages within a basin, and also, between basins.

The usefulness of the variability ratio is illustrated in the comparison of the Manistee River and the Raisin River. The variability ratio for the Manistee River is about 0.9, which is to say that the 1-in-10-year drought flow is 90 percent of the most probable, indicating an unusually stable stream. In contrast, the Raisin River record develops a variability ratio of about 0.3, which is to say that the 1-in-10-year drought is only 30 per cent of that normally expected, indicating a river of high variability from year to year and subject to occasional drought flows of considerable severity.

BASIN SUMMARY

In many of the basins of the State where three or more representative gages existed, it has been possible to establish a linear relationship between the logarithm of the drainage area size and the logarithm of the minimum consecutive 30-day average most probable and 1-in-10-year droughts. Such a relationship for the Kalamazoo River is illustrated in Figure 13 on a log-log scale, with Curve A the most probable drought, and Curve B, the 1-in-10-year drought.

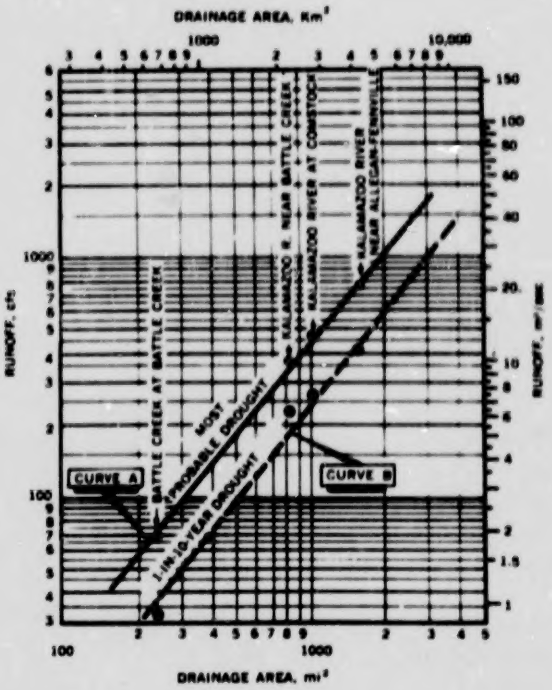


Figure 13 — Summer-Fall Drought Flow as Minimum in Kalamazoo Basin, Consecutive 30-Day Average versus Tributary Drainage Area.

In addition to serving as a summary for the key gages in the basin, this chart is useful in estimating drought flows along the river at points that do not have a stream gage, but where the tributary drainage area is known. For example, on the Kalamazoo River at a point having a drainage area of 700 square miles, the most probable 30-day average drought would be estimated from Curve A as 280 cfs and 1-in-10-year average drought from Curve B as 150 cfs.

NATURAL AND ARTIFICIAL INFLUENCES

In dealing with drought flows, the investigator must be continually alert to the possibility of either natural or artificial influences. Natural influences might be reflected in a drainage basin with widely varying yield characteristics, whereas artificial influences might include many of man's activities such as hydroelectric and steam power production, diversion for irrigation or municipal or industrial use, or possibly navigation or even flood protection facilities.

A good example of a river with widely varying yield characteristics is the Willamette River in Oregon, especially in the section of the river from Salem to Portland. Tributaries on the eastern side fed by the melting snows of the Cascades produce high yields, whereas those from the Coastal Range on the western side produce low yields. Figure 14 illustrates an attempt to estimate the once-in-5-year minimum weekly average flow at Portland by considering the yield at Salem together with the yields from the individual tributaries. It will be noted that the Yamhill has a yield of 0.052 cfs per square mile, whereas the Clackamas draining the Mt. Hood area has a yield of 0.763 cfs per square mile. The figures used in Figure 14 represent flow conditions prevailing in the Willamette

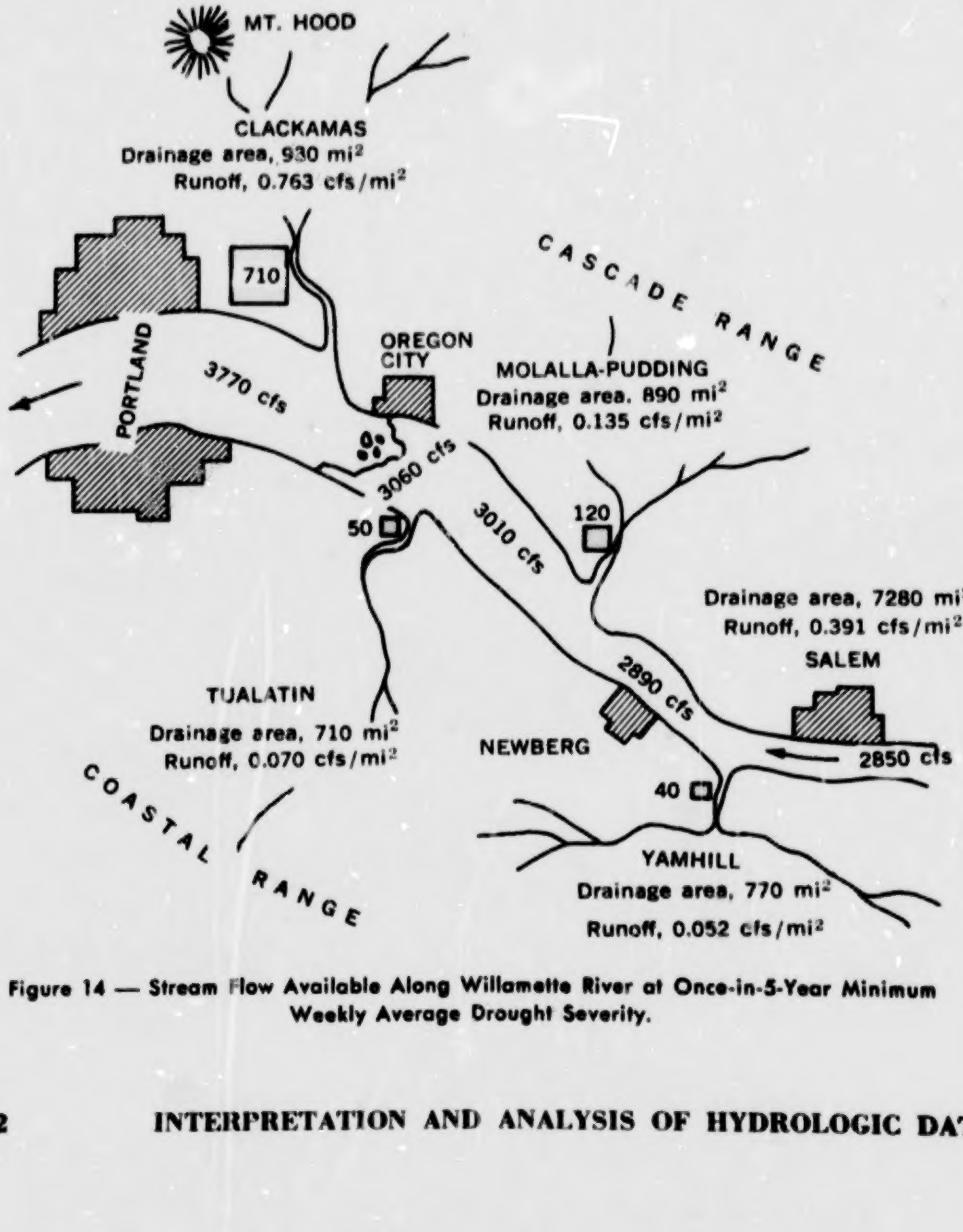

MT. HOOD
CLACKAMAS
Drainage area, 930 mi^2
Runoff, 0.763 cfs/mi^2
710
PORTLAND
3770 cfs
OREGON CITY
CASCADE RANGE
MOLALLA-PUDDING
Drainage area, 890 mi^2
Runoff, 0.135 cfs/mi^2
3060 cfs
50
3010 cfs
120
Drainage area, 7280 mi^2
Runoff, 0.391 cfs/mi^2
SALEM
2890 cfs
TUALATIN
Drainage area, 710 mi^2
Runoff, 0.070 cfs/mi^2
NEWBERG
2850 cfs
40
COASTAL RANGE
YAMHILL
Drainage area, 770 mi^2
Runoff, 0.052 cfs/mi^2

Figure 14 — Stream Flow Available Along Willamette River at Once-in-5-Year Minimum Weekly Average Drought Severity.

prior to 1950 and are not illustrative of present low flows in the main river, which are influenced by low-flow augmentation resulting from upstream storage. Notwithstanding these changes, the illustration does indicate the dramatic differences in yield from the tributaries on the eastern and western side of this section of the river.

Under the category of an artificial influence might be considered the diversion of river water around the Geological Survey stream gage on the Tittabawassee River at Midland by the Dow Chemical Company for industrial use. This diverted flow is returned to the river below the gage, together with a small amount of imported Lake Huron water, resulting in an augmentation of the natural river drought flows. The influence of the diversion was illustrated during a special time-of-passage study conducted by the writer and his associates under controlled river flow conditions on August 15, 1962. Table 1 tabulates the Geological Survey stream gage flow, together with the flows not reflected by this gage but returned to the river downstream from the gage.

Table 1 — River Flows in Tittabawassee River Below Midland, Michigan, August 15, 1962.

Geological Survey stream gage	220 cfs
Dow treatment plant effluent	76.2 cfs
Drain A	13.4 cfs
Drain B	5.3 cfs
Total river flow below Dow	314.9 cfs

It is interesting that a discharge measurement of the Tittabawassee River taken independently by the writer at the first convenient sampling station downstream from the diversion amounted to 314.7 cfs, indicating excellent agreement with the sum of the individual upstream measurements. If reliance were placed only on the official Geological Survey flow measurement as an indication of downstream flow in the Tittabawassee River during this survey period, the estimates would be in serious error.

FLOW REGULATION

In many river basins, low-flow regulation can be accomplished by storing high flows and releasing them during the dry-weather period of the year, thereby eliminating the most severe drought conditions. As is generally known, one of the most important elements governing the waste assimilation capacity of a stream is the flow level; the higher the flow, the greater the capacity; the lower the flow, the less the capacity. Thus, flow regulation eliminates the need for controlling waste discharges to meet water quality needs under the most severe drought conditions, and in many cases, makes it possible to work with guaranteed minimum flows substantially greater than the natural dry-weather flows. Where storage is in the headwater of a stream, the benefits accrue not only to the section of the river involving waste assimilation, but to all other sections of the river below the impoundment, including water for municipal and industrial water supply, power production, and recreation.

One of the important considerations, of course, is the availability of a suitable site or sites for reservoir development. A systematic study of the headwater and downstream tributaries may yield several locations that could be used as reservoir sites, and therefore, would merit further analysis. Several years ago such a study was conducted in the Kalamazoo River basin by Velz and Gannon,[10] resulting in the location of a favorable reservoir site on one of the upstream tributaries.

Figure 15 shows a comparison of the potential regulated flow and the natural 10-year drought flow (weekly and daily averages) for the Kalamazoo River at Kalamazoo, at Battle Creek, and at Marshall for the key dry-weather months of July through October. It is apparent that at Kalamazoo a regulated flow of approximately 700 cfs could be maintained in comparison to the natural drought flows in the vicinity of 200 cfs or less. In addition, it is seen that not only does the river in the vicinity of Kalamazoo benefit, but also, there is a substantial increase over natural drought flows at Battle Creek and at Marshall.

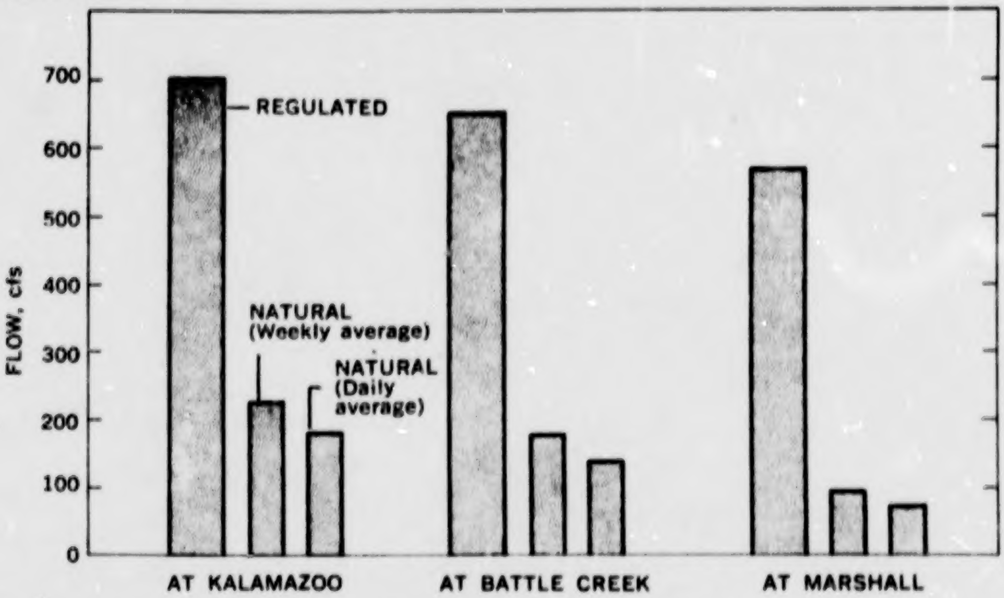


Figure 15 — Comparison of Potential Regulated Flow and Natural 10-Year Drought Flow in Kalamazoo River.

A major pollution problem exists on the Kalamazoo River below the City of Kalamazoo, and one of the main benefits of increased low flows would be the improvement of water quality. Unfortunately, an economic study made subsequently by another group indicated that it would be less costly to improve water quality through this section by additional waste treatment rather than by means of flow augmentation. As a result, the proposal for low-flow augmentation by reservoir development was not considered further and major reliance for water quality improvement is being placed on additional waste treatment. It may be that as the demand for water increases in the future the economic balance will change and this proposal will receive further consideration.

Standard procedures for the determination of storage needs by mass curve analysis, etc., is covered adequately in such text books as that of Fair and Geyer,[11] and will not be considered here.

TIME OF PASSAGE

One of the important elements necessary for an accurate evaluation of the self-purification capacity of a river where one or more different types of wastes, such as organic, bacteriological, or chemical contaminants, may exist is the time of flow or passage along the stream. This information may be obtained or estimated in several ways. From a knowledge of the channel characteristics and prevailing runoff and use of an internal or external tracer such as a dye, it can be calculated on a displacement basis, or from a knowledge of certain generalized data, it can be estimated as proposed by O'Connor.[12]

DISPLACEMENT CALCULATIONS

Where the river channel is of a fairly uniform character, time of passage can be calculated for a given runoff on a displacement basis. This presumes that information is available on the river channel characteristics so that accurate volumes can be calculated for a given runoff level. Sometimes this information is available from sources such as the files of Corps of Engineers units covering flood protection or making navigation studies. It may be necessary to collect this information in the field; if this is the case, adequate definition of channel characteristics can generally be obtained by cross-sectioning the relevant river stretches at about 500-foot intervals. This need not be done with a high degree of accuracy, but rather emphasis should be placed on more frequent soundings wherever possible. It may be accomplished by means of a tape and sounding rod or weighted line, together with a good map for location and orientation in the field, or if considerable cross-section work is anticipated, it might be desirable to obtain a portable recording fathometer that gives a continuous record of channel depth.

Volumes can be calculated on an average end-area basis with adjustment to various runoff levels made by means of an appropriate rating curve. This approach can be programmed for high-speed digital computers and incorporated as a part of a more extensive program such as that described by Gannon and Downs[13] for programming river dissolved oxygen calculations.

Figure 16 contains a series of time of passage curves calculated on the displacement basis for various runoff levels for the Willamette River, for the section of the river extending from Salem to Oregon City Falls. Fortunately, in this case, detailed charts with frequent depth figures were available from the Corps of Engineers[14] and these data served as the basis for volume calculations.

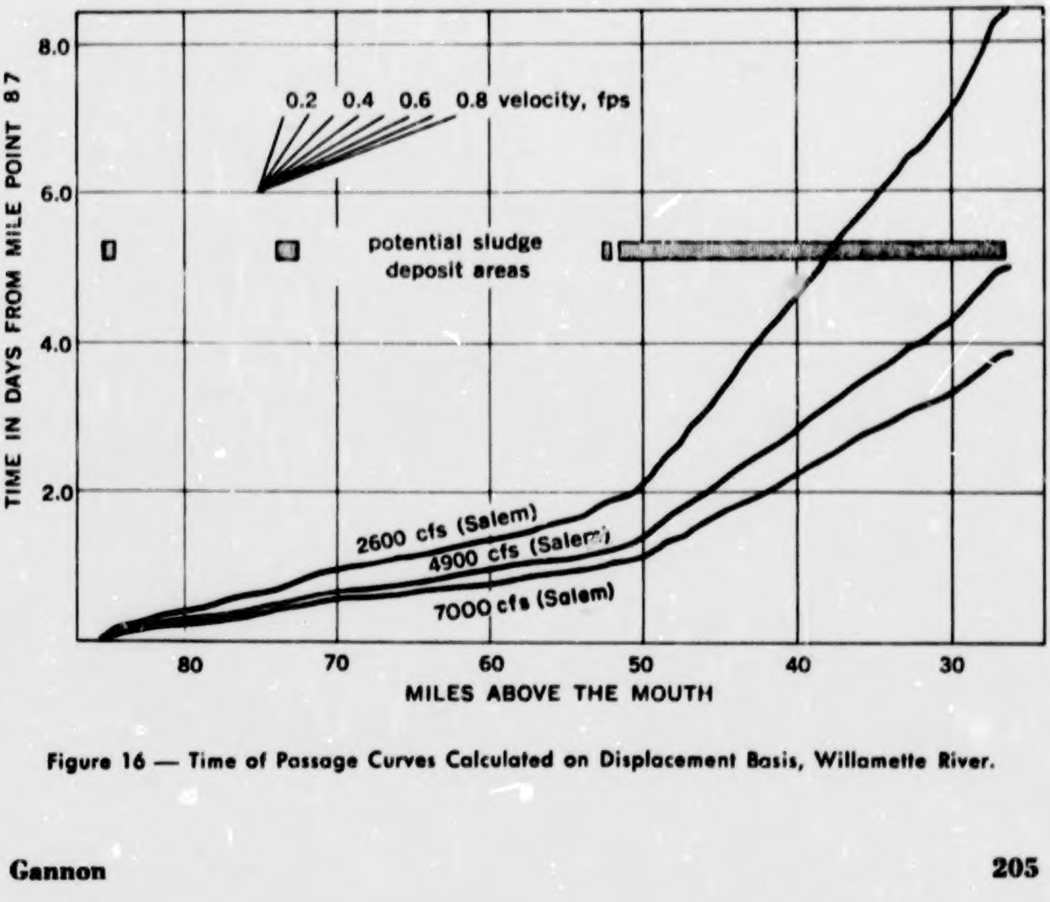


Figure 16 — Time of Passage Curves Calculated on Displacement Basis, Willamette River.

In Figure 16 the slope of the time curve is in effect a measure of the average velocity in that section of the river. This type of plot, therefore, serves as an excellent guide in identifying those sections of the river that would serve as potential sludge deposit areas. Velz[15, 16] has indicated that at velocities of 0.6 fps or less settleable solids deposit and tend to accumulate to an equilibrium level. Thus, the nest of velocity curves in Figure 16 show that a potential sludge deposit area exists in the pool section of the river extending from approximately mile point 52 to 26, and also, in short sections in the stretch from mile point 87 to 52. If, therefore, any wastes containing settleable solids were discharged into the river above these potential deposit areas, it is almost certain that sludge deposits would develop, resulting in oxygen deficient conditions.

Knowledge of the channel characteristics is necessary for other purposes, such as the calculation of reaeration for oxygen balance needs where an organic waste problem exists. Here, it is important to know both mean depths and volumes for the critical section of the river.

INTERNAL AND EXTERNAL TRACERS

In addition to the displacement approach, time of passage can also be determined by means of either internal or external tracers. An internal tracer may be classified as some waste constituent that can be varied in concentration and easily measured in the stream, e.g., chlorides. An external tracer may be classified as anything that can be added to the stream and then easily followed and measured, e.g., salt, dye, and radioactive material. Several investigators have reported on tracer methodology, including Carpenter[17] on Chesapeake Bay, Selleck and Pearson[18] on San Francisco Bay, and Hull[19] on the American River in California.

As part of recent investigations on the Tittabawassee River, an opportunity developed for making comparisons of times of passage (1) calculated on the basis of displacement, (2) use of waste chloride concentration as an internal tracer, and (3) use of Rhodamine B fluorescent dye as an external tracer, together with a sensitive fluorometer for detection purposes.

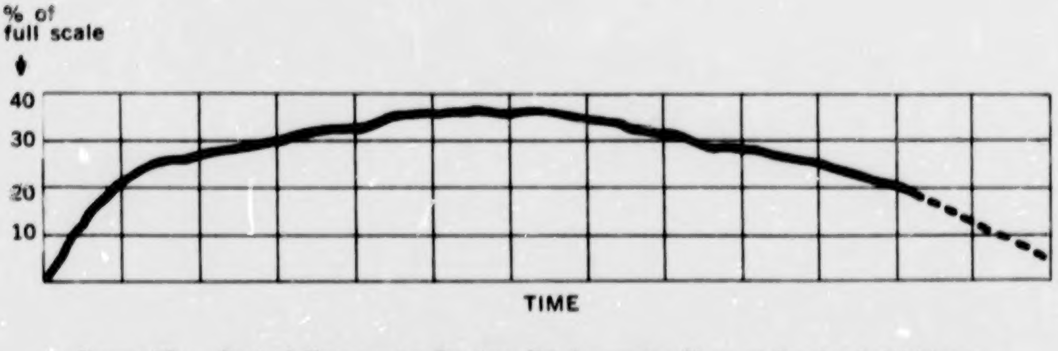


Figure 17 — Typical Fluorometer Tracing, Tittabawassee River at Freeland, Michigan, August 15, 1962.

Figure 17 is a typical fluorometer tracing of Rhodamine B dye detected at Freeland, on the Tittabawassee River on August 15, 1962, during a controlled river flow condition for special time of passage studies. The dye was introduced as a point discharge at the next upstream station; as a result of longitudinal mixing or dispersion, it took the dye approximately 3.5 hours to pass the station at Freeland. For a more complete discussion of the mixing and diffusion of wastes in streams, the reader is referred to the work of Thomas.[20]

Figure 18 is a comparison of calculated and observed times of passage on the Tittabawassee River below Midland for a total river flow in the range of 315 to 350 cfs. The calculated times were determined on the displacement basis; however, the channel was cross-sectioned under higher runoff levels and it was necessary to adjust the volume down to the indicated runoff levels. The observed flow time was reported by the Dow Chemical Company on the basis of chloride concentration studies conducted many years ago. The dye tracer studies were directed by the writer under controlled river conditions on August 15, 16, 1962.

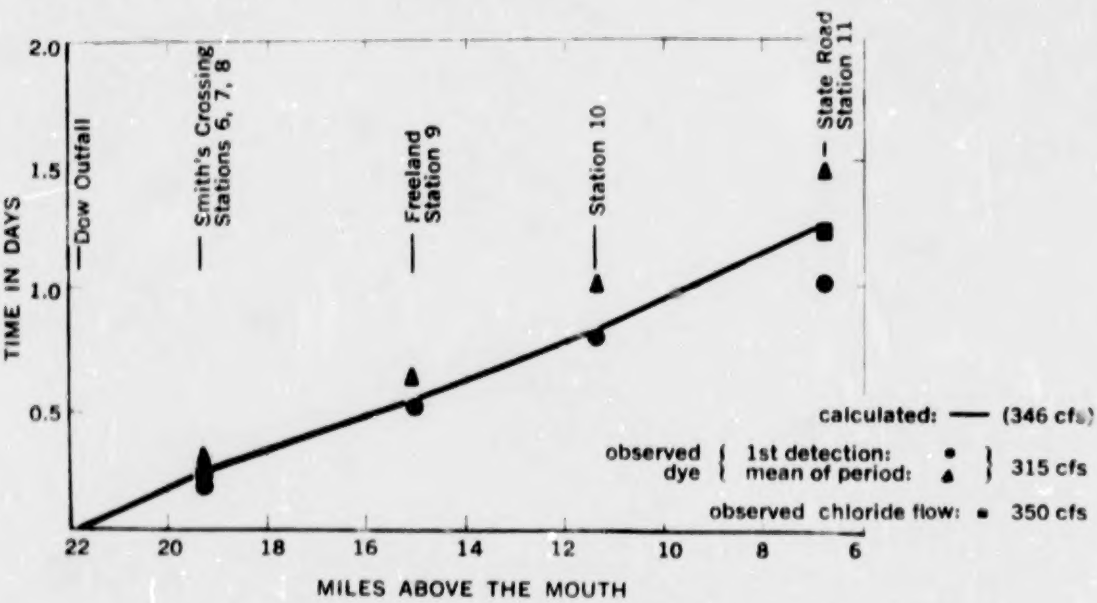


Figure 18 — Comparison of Calculated and Observed Times of Passage, Tittabawassee River.

The results of the Rhodamine B study are plotted in two ways: first, as the time of first detection, which might be important if toxic wastes were involved, and second, as the mean of the period, which should be compared with the displacement calculation. Reasonable agreement exists to Freeland, but differences are greater farther downstream. The dye studies were conducted at a runoff level of 315 cfs; the calculated time corresponding to a runoff of 346 may partially account for the differences. Furthermore, a backwater influence at station 11 from Lake Huron no doubt contributes to the differences at this station. The river channel is fairly uniform and shallow through this section, suggesting minimum amounts of short circuiting.

The availability of dyes such as Rhodamine B and Pontacyl Brilliant Pink B, together with sensitive fluorometer detection instruments, makes an external tracer study of time of passage a necessary part of any well-planned stream survey.

ACKNOWLEDGMENTS

The assistance of Mr. Jackson R. Pelton in the time of passage studies, and Mrs. Josephine Toney in the statistical compilations and computations is gratefully acknowledged.

Financial support was provided by the Water Resources Commission of the State of Michigan for the analysis of drought flows in Michigan, while the U. S. Public Health Service supported studies on the Clinton and Tittabawassee Rivers as research grant RG-6905 later redesignated WP-187.

The cooperation of Mr. Arlington Ash, District Engineer, U. S. Geological Survey, in supplying runoff information, and also, the cooperation of Messrs. John Robertson and Charles Sercu of the Waste Control and Utilization Department, Dow Chemical Company, Midland, Michigan, in supplying information and facilitating several phases of the Tittabawassee study is gratefully acknowledged.

Finally, particular recognition is due to Professor C. J. Velz, Chairman of the Environmental Health Department, The University of Michigan, who has encouraged the use of the statistical tool in the analysis of hydrologic data, particularly as it relates to water quality considerations.

REFERENCES

1. "Water Supply Papers." Published annually by U. S. Geological Survey.
2. Hazen, Allen, "Storage to be Provided in Impounding Reservoirs for Municipal Water Supply." *Transactions of the American Society of Civil Engineers*, 77, 1539 (1914).
3. Velz, C. J., "Graphical Approach to Statistics." *Water and Sewage Works*, 99, 4, R106 (1952).
4. Gannon, John J., "Statistical Basis for Interpretation of Data." *Proceedings of Michigan Sewage and Industrial Wastes Association 1959 Annual Meeting*, 34 pp. (1959).
5. Velz, C. J. and Gannon, John J., "Drought Flow of Michigan Streams." Michigan Water Resources Commission, Lansing, 771 pp. (1960).
6. Gumbel, E. J., "Statistical Theory of Droughts." *Proceedings American Society of Civil Engineers*, 80, separate No. 439 (May, 1954).
7. Gumbel, E. J., "Statistical Theory of Floods and Droughts." *Journal of the Institution of Water Engineers* (British), 12, 3, 157 (May, 1958).
8. Gumbel, E. J., "Statistics of Extremes." Columbia University Press, New York (1958).
9. Velz, C. J. and Gannon, John J., "Low Flow Characteristics of Streams." *Proceedings of the Second Annual Ohio Water Clinic*, Ohio State University Studies Engineering Series, 22, 4, 138 (1953).
10. Velz, C. J. and Gannon, John J., "Reservoir Site Study in the Kalamazoo Basin." Unpublished material.
11. Fair, Gordon M. and Geyer, John C., "Water Supply and Waste-Water Disposal." John Wiley and Sons, Inc., New York (1954).
12. O'Connor, Donald J., "The Effect of Stream Flow on Waste Assimilation Capacity." Paper presented at the 17th Purdue Industrial Waste Conference (May, 1962).
13. Gannon, John J. and Downs, Thomas D., "Programming River D. O. Calculations." *Water and Sewage Works*, Part I, 110, 3, 114 (March, 1963); Part II, 110, 4, 157 (April, 1963).

14. U. S. Corps of Engineers, Willamette River, Oregon. Portland, Oregon Office (Revised to November, 1938).

15. Velz, C. J., "Factors Influencing Self-Purification and Their Relation to Pollution Abatement — Part II — Sludge Deposits and Drought Probabilities." *Sewage and Industrial Wastes*, 21, 2, 309 (1949).

16. Velz, C. J., "Significance of Organic Sludge Deposits." *Oxygen Relationships in Streams*, Technical Report W-58-2, Robert A. Taft Sanitary Engineering Center (1958).

17. Carpenter, James H., "Tracer for Circulation and Mixing in Natural Waters." *Public Works*, p. 110 (June, 1960).

18. Selleck, Robert E. and Pearson, Erman A., "Tracer Studies and Pollutional Analyses of Estuaries." Publication No. 23, State of California Water Pollution Control Board, Sacramento (1961).

19. Hull, D. E., "Dispersion and Persistence of Tracer in River Flow Measurements." *International Journal of Applied Radiation and Isotopes*, Vol. 12, p. 63 (1961).

20. Thomas, Jr., Harold A., "Mixing and Diffusion of Wastes in Streams." *Oxygen Relationships in Streams*, Technical Report W-58-2, Robert A. Taft Sanitary Engineering Center (1958).

DISCUSSION

Mr. Gray asked whether observations of dye concentration had been carried on for any period of time after the low point apparently was reached. He indicated that a second peak had been observed in a test in which grab sampling had necessitated an extended sampling period to assure that the dye had passed.

Mr. Gannon observed that this is probably the result of pools in the stream that hold some of the dye and then feed it back to the river. The dye could not be followed for any great length of time in the Titabawassee River, since the flow was being reduced by storage in a reservoir of limited capacity. The investigators observe the time of first detection and the time of maximum concentration, and estimate the remainder of the curve. He believes that dye provides a relatively simple, inexpensive means of securing fairly accurate estimates of time of passage.

Mr. O'Connell noted that the time of first appearance is somewhat short of the calculated displacement time, while the mean time is a little longer than the calculated time. He asked whether the time of passage as indicated by the peak concentration had been considered, since it might be closer to the displacement time. The mean might give a distorted measure of time of passage because of the diffusion that takes place while the dye passes the measuring station. A synoptic observation would prevent distortion from this source.

Mr. Gannon agreed that the peak is sometimes used rather than the mean to determine time of passage. He indicated that diffusion prevents the following of a slug of dye very far downstream. Observing dye passage from one station to the next is the most practical method, and problems of background concentrations are avoided by starting with the downstream station and working up river with a new slug of dye each time.

SESSION 6: General

Chairman: **Bernard B. Berger**
Assistant Chief for Research
Division of Water Supply and Pollution Control
U. S. Public Health Service

Dr. John C. Bellamy
Director, Natural Resources Research Institute
University of Wyoming, Laramie

SUMMARY

The purpose of informatic data research is to find better ways of using new equipment that make it possible to acquire and analyze vast amounts of quantitative data. The kinds of data of primary interest are those that will better inform us about the nature of the ground, water, air, and near-space environment of our geosphere. New forms of numerals can advantageously be utilized to paint half-tone pictures that will not only provide the qualitative information needed to gain understanding of our environment, but can also serve as a concise and complete data store for whatever arithmetic processing anyone might subsequently wish to have performed by machine. This is one of a series of papers describing the progress of Informatic Data Research at the University of Wyoming.

DATA DISPLAY FOR ANALYSIS

This is one of a series of papers describing the progress of a program of Informatic Data Research at the University of Wyoming. The purpose of this program is to establish the principles and practices of utilizing newly possible

informatic	ways of representing large sequences of numbers as concise complete "pictures" or portrayals of
information	which can be acquired, processed, recorded, and reprocessed in numerical detail only with appropriate
automatic	equipment if scientific and engineering operations are to become more economically effective.

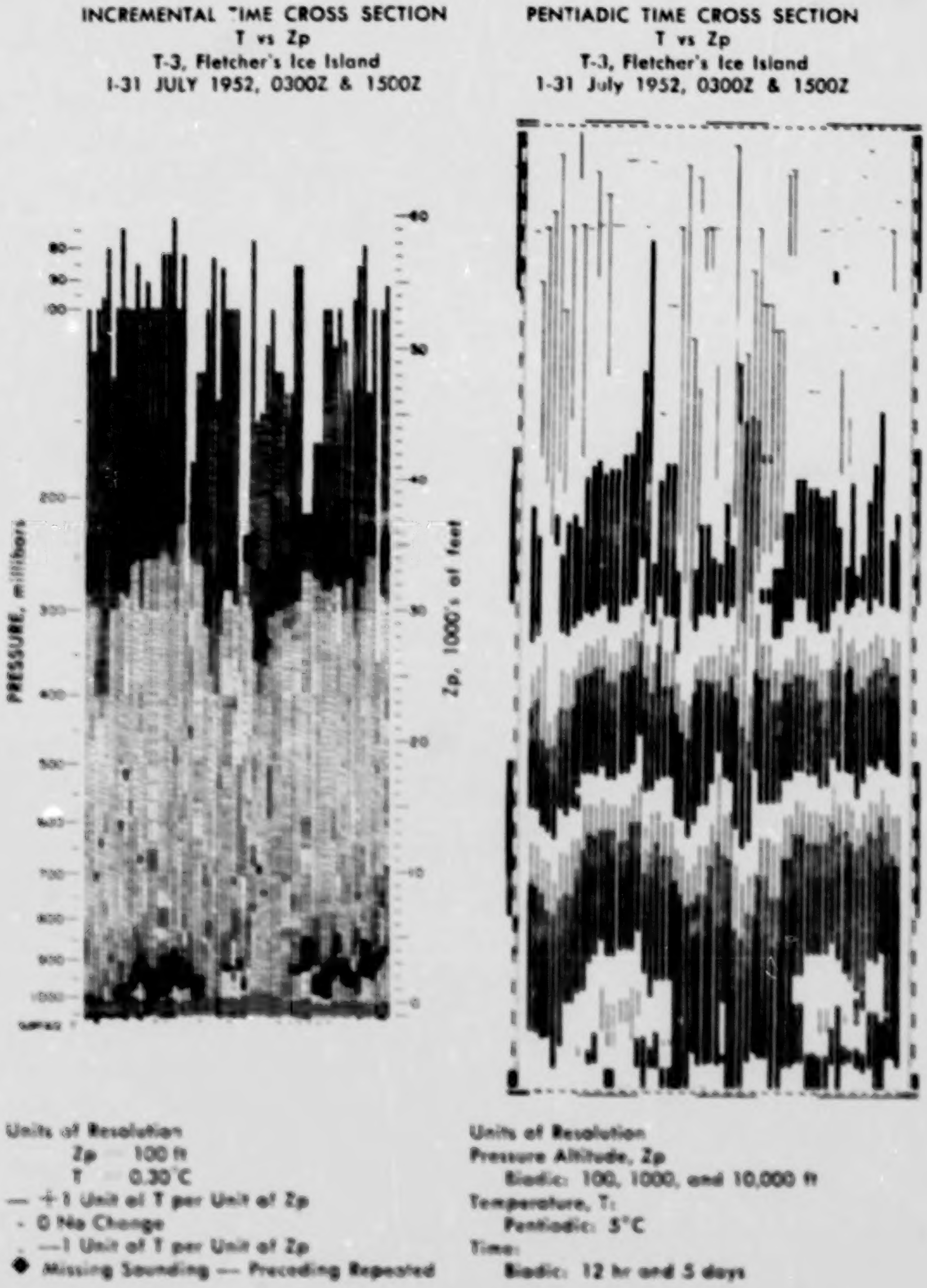
In brief review of a previous discussion,[1] the goal of informatic data research is to find ways of utilizing newly possible equipment for better acquiring, analyzing, and utilizing vast amounts of quantitative data. The kinds of data of primary interest are those that will better inform us about the nature of our ground, water, air, and near-space environment, or in short, about the nature of the lithosphere, hydrosphere, atmosphere, and pyrosphere that make up the geosphere.[2]

Progress to date has shown that new forms of numerals can advantageously be utilized now that man no longer has to write them. Briefly, these new numerals can be likened to the variably sized dots that make up half-tone reproductions of photographs or paintings. Consequently, the goal of informatic data research can be thought of as being to find ways of utilizing machines "to paint half-tone pictures with numerical data."

Even partial realization of this goal would evidently be very worthwhile. Not only would the resultant "pictures" provide men with the largely qualitative kind of information they need to gain understanding of the nature of their environment, these "pictures" would also serve as a concise and complete data store for whatever arithmetic processing anyone might subsequently wish a machine to perform. The sizes of the individual "half-tone numerals" would need only to be sensed by appropriate reading equipment to re-establish whatever numerical values might be needed for any desired quantitative analysis.

INCREMENTAL DATA

To illustrate, the data-block on the left in Figure 1 is an "incremental" portrayal of the vertical distributions of values of temperature measured throughout a month above a particular radiosonde station.[3]



INCREMENTAL TIME CROSS SECTION
T vs Zp
T-3, Fletcher's Ice Island
1-31 JULY 1952, 0300Z & 1500Z

PENTIADIC TIME CROSS SECTION
T vs Zp
T-3, Fletcher's Ice Island
1-31 July 1952, 0300Z & 1500Z

Units of Resolution
Zp — 100 ft
T — 0.30°C
— +1 Unit of T per Unit of Zp
· 0 No Change
. —1 Unit of T per Unit of Zp
◆ Missing Sounding — Preceding Repeated

Units of Resolution
Pressure Altitude, Zp
Biodic: 100, 1000, and 10,000 ft
Temperature, T:
Pentiodic: 5°C
Time:
Biodic: 12 hr and 5 days

Figure 1 — Incremental and Iodic Notations.

To understand the pictorial character of this example, it can be thought of as though it had been produced as a half-tone reproduction of a photograph of a plaster-of-paris model of that measured temperature distribution. The height of each point of the model would have been proportional to the value of temperature measured at the corresponding values of altitude (or pressure) and time, and it would have been photographed with a point source of illumination above and to the high-altitude side of the model. The light regions would then have occurred where the illumination was normally incident upon the surface of the model, or where the temperature increased rapidly downward; gray regions would have occurred where the illumination was obliquely incident, or where the temperature was nearly constant vertically; and dark regions would have occurred in regions of grazing illumination, or where the temperature increased rapidly upwards. Or, in meteorological terms, the troposphere is light, the stratosphere is gray, the low-level and tropopause inversions are black, and the various shades of gray indicate various degrees of lapse rate.

Actually, of course, this incremental data block was produced without going to the trouble of constructing and photographing a plaster-of-paris model. Rather, it was formed in accordance with a particularly simple arithmetic formula based upon the characteristics of continuous data. That is, in order to explicitly represent the value of temperature measured at each and every altitude, it is necessary that no significant changes of temperature be omitted and hence that the data be continuous in the sense that no two successive values differ by more than an appropriately selected and significant unit of numerical resolution. But then, since the numerical differences, or increments, between successive values can only be +1, 0, or —1, it is only necessary to record one of these three possible values of increments between each successive value of continuous data to designate all but its initial values.

Specifically, units of resolution of 1/3°C, 100 feet, and 12 hours are used for temperature, (pressure) altitude, and time in this example. Incremental numerals consisting of short, medium, and long horizontal dashes have been used (instead of the Arabic numerals —1, 0, and +1) to represent, respectively, a unit of 1/3°C decrease, no decrease or increase, and a unit of 1/3°C increase of temperature over a unit increase of altitude. Initial or ground level values of temperature for each sounding have been tallied with similar numerals at the bottom of the data block.

In other words, more than 36,000 measurements of temperature obtained at about 600 increments of altitude in each of 62 radiosonde soundings are contained in this example. Any or all such values of temperature could readily be re-established with an optical sensor that need only identify three widths of marks while scanning any one of the sounding-data lines from bottom to top. Counting the marks without regard to their size would provide values of altitudes, and counting the marks with regard to their size or algebraic sign would provide corresponding values of temperature.

IADIC DATA

Although such an incremental form of data provides most of the desired characteristics of conciseness, qualitative portrayal, and quantitative exactness, it falls short of the ideal in two respects. First, it is virtually impossible to discern particular quantitative values manually. Second, it is too nonredundant; any error in sensing an incremental value would produce a continuing and undectectable error in all succeeding counts of "whole" values.

An early attempt to eliminate these shortcomings is illustrated on the right in

Figure 1. These data are recorded with an "iadic" or "*i*ncrementally *a*lternating *d*ash *i*ncrementally *c*ontinuous" notation. In effect, numerals in the iadic notation consist of marks with variable transverse widths, each width standing for the value (such as zero, one, two, three, or four in this example) of some particular digit in a digital representation of a "whole" number. The name of the notation is derived from the fact that a long longitudinal "dash" is formed by the continual repetition of a particular width of numeral throughout a region in which the value of the digit remains constant, and that a one-to-one correspondence exists between a change of value of the digit and a change of width of the longitudinal "dashes."

In this particular example, a "pentiadic" notation has been used to record the number of 5°C units contained in the "whole" values of temperature represented incrementally on the left of the page. That is: a "zero width" of iadic marks designates that the temperature lies between 0° and +5°, —25° and —20°, —50° and —45°; a "one width" of mark designates a temperature between 5° and 10°, —20° and —15°, —45° and —40°; and so on until a "four width" of mark designates a temperature between +20° and +25°, —5° and 0°, —30° and —25°, and between —55° and —50°. Consequently, the positions of particular isotherms at 5° intervals are readily apparent as the positions at which the width of the dashes change. The positions of isotherms at 25° intervals (or at 0°, —25°, and —50°) are especially apparent as being the positions at which the width changes between the "zero width" and the "four width."

The data display characteristics of the iadic form of notation can thus be summarized as follows. It provides a readily apparent "picture" of the large scale distributions of particular quantitative values in much the same way as isotherm or contour maps do. It would also provide for error-checking in automatic playback of associated incremental records. In that case, corresponding iadic and incremental data lines (or soundings) would be scanned simultaneously. They would then be checked to see that each change in width of the iadic numeral corresponds with a change of value of that digit in a continuing count of the incremental changes.

TALLIC DATA

The two examples in Figure 1 suffer the disadvantage that although they should be used together it is very difficult to use them together. For example, it would be virtually impossible to maintain the degree of mechanical registry required for simultaneous error-checking scanning of the two data-blocks as they appear in Figure 1. On the other hand, if corresponding incremental and iadic sounding lines were to have been interspersed into juxtiposition, most of the highly desirable shades and shadows pictorialization would have been lost.

To overcome this disadvantage, several attempts have recently been made to utilize a "tallic" or "*t*ransversely *a*nd *l*ongitudinally *l*abelled, *i*ncrementally *c*ontinuous," or "tally-like," form of numerals. That is, it is evidently possible to vary both the transverse width and longitudinal thickness of rectagular "half-tone dots" in order to simultaneously tally two kinds of interrelated numbers.

The first and most readily accomplished trial[4, 5] resulted in the formulation of the tallic notation illustrated in Figure 2. The goal was to represent as concisely and clearly as possible those years for which observational data were obtained at some particular station with instruments such as stream gages or precipitation gages. This goal was realized by using four transverse widths of tally marks to designate groups of 5 years each in a repetitive 20-year pattern. Also the availability of full, partial, or no record of measurement for a particular year is indicated, respectively, by a longitudinally thick, medium, or thin tally mark for that year. The utility of this kind of notation is indicated by the combination of over-all view and copius quantitative detail portrayed by the enclosed map (Figure 3) of the periods of records available from all stream gage stations that have ever existed in Wyoming since 1890.

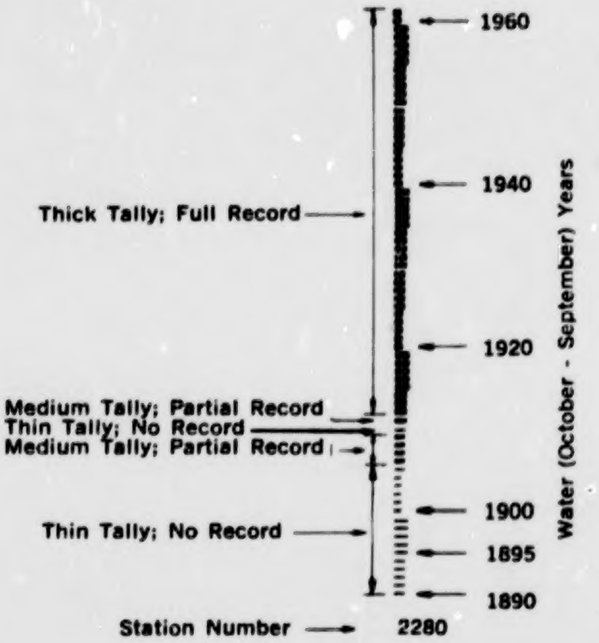


Figure 2 — Sample Tallic Representation of Periods of Data Records for the Period from 1890 through 1961.

The results of a more ambitious attempt[5] to develop and use a tallic notation is illustrated in Figure 4. This particular example portrays (1) each of the hours throughout 4 years in which at least 0.01 inch of precipitation fell on the precipitation gage at Laramie, and (2) the running accumulation of precipitation throughout each of those years.

Briefly, the periods and rates of precipitation are indicated in the following way. A row of 365 (or 366) lines, each consisting of 24 side-by-side tally marks, identify each hour of each day of each year. The occurrence of at least 0.01 inch of precipitation during any particular hour is indicated by a thick tally mark for that hour. A medium-thick tally mark is used to indicate that it did not precipitate during that hour, but that a precipitation amount in excess of 0.01 inch had occured (and had not yet been accounted for) in some closely preceding hour. A thin hourly tally mark indicates (1) that no precipitation fell during that hour and (2) that the total number of preceding thick and medium-thick tally marks equals the total number of hundreths of inches of precipitation that had previously fallen that year.

In addition, accumulated amounts of precipitation are indicated by five different transverse widths of the hourly tally marks. The narrowest width of mark is used to indicate accumulations between 0.0 and 0.2 inch, 1.0 and 1.2 inches, 2.0 and 2.2 inches, etc.; the next wider width of mark indicates accumulations between 0.2 and 0.4 inch,

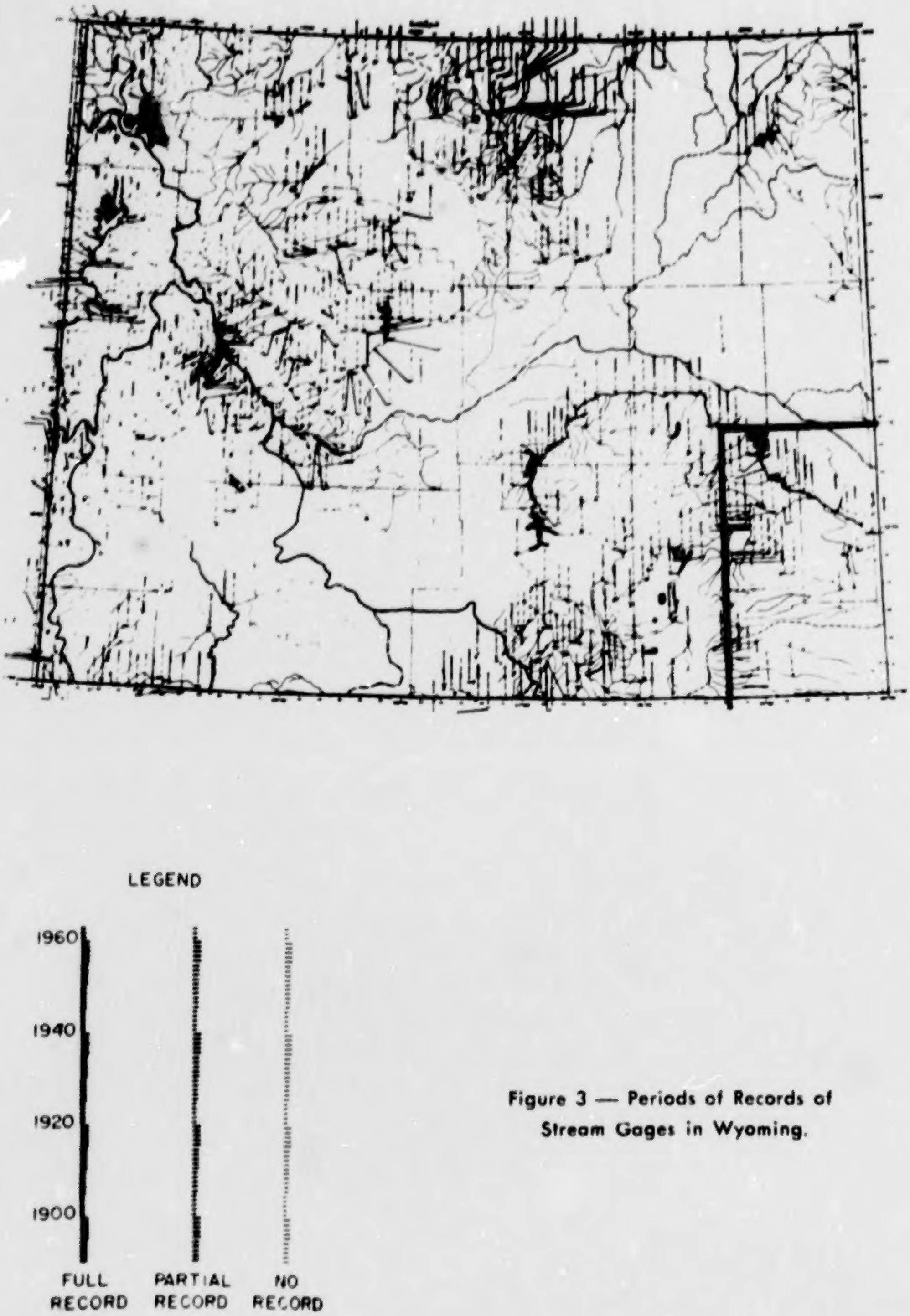
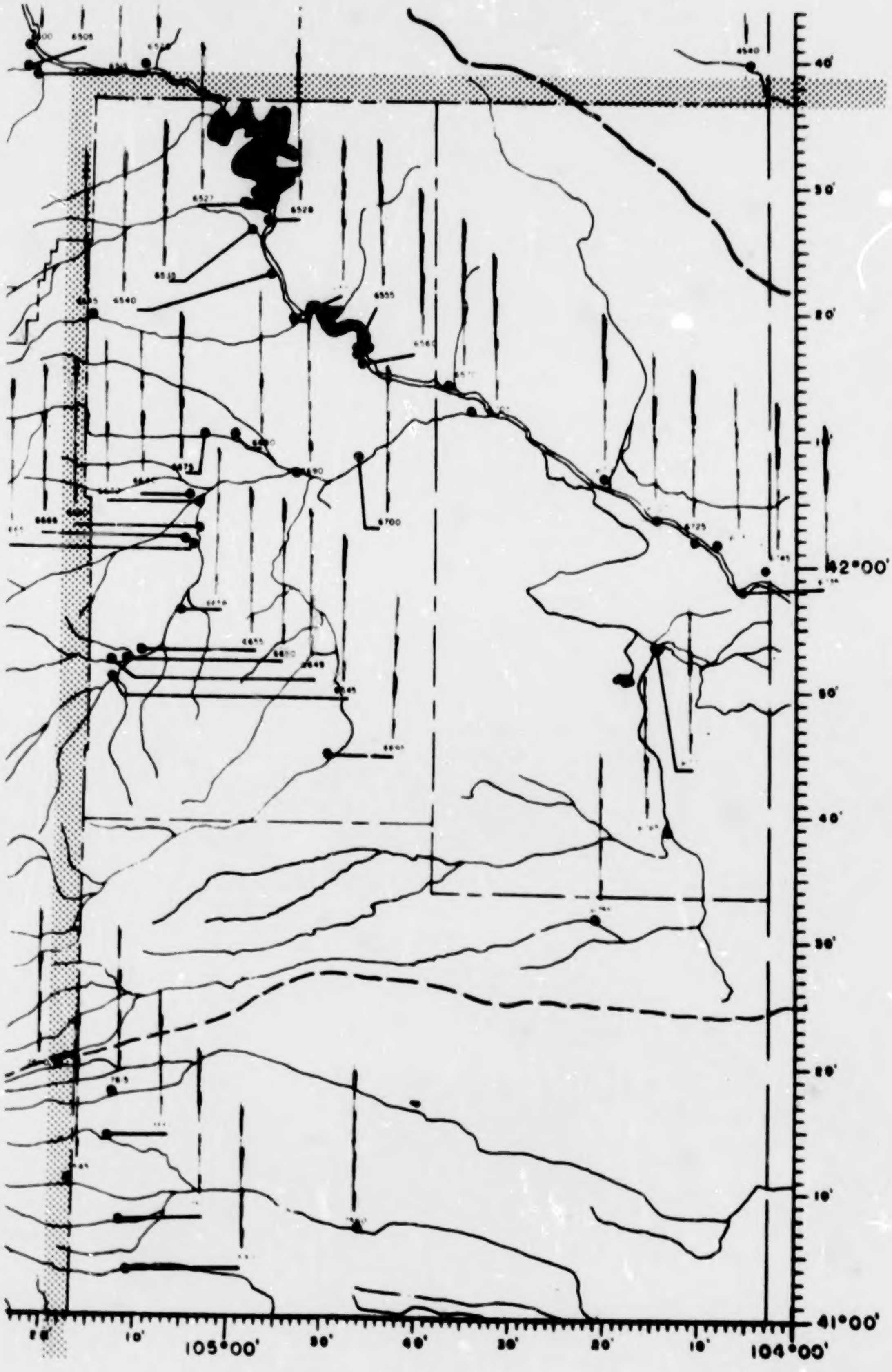

LEGEND
1960
1940
1920
1900
FULL RECORD
PARTIAL RECORD
NO RECORD

Figure 3 — Periods of Records of Stream Gages in Wyoming.

GPO 814-105-8


6500
6505
6520
6540
6527
6528
6510
6535
6540
6555
6580
6700
6725
40'
30'
20'
10'
42°00'
50'
40'
30'
20'
10'
41°00'
20'
10'
105°00'
50'
40'
30'
20'
10'
104°00'

1.2 and 1.4 inches, 2.2 and 2.4 inches, etc.; and the widest of the five widths of mark indicates accumulations between 0.8 and 1.0 inch, 1.8 and 2.0 inches, 2.8 and 3.0 inches, etc. The gray shade appearance of this particular example is determined primarily by these transverse widths of the tally marks. Hence, it is relatively easy to determine the full inches of accumulation by counting the number of times they have changed from their widest and darkest appearing width to their narrowest and lightest appearing width.

The results of a similar attempt[6] to develop and use a tallic notation for portraying rates of flow of three streams throughout 10 years are reproduced in Figure 5. In each of these examples, a line of 365 (or 366) tally marks is used to identify each day of each year. Three different longitudinal thicknesses of tally marks are used to indicate (1) that the flow increased to at least one appropriately selected unit of measurement of flow more than on the previous day with a thick tally mark, (2) that the flow decreased to at least one unit of measurement less than on the previous day with a thin tally mark, or (3) that the flow remained constant within these limits with a medium-thick tally mark. Five transverse widths of tally marks are used to indicate that the rate of flow on any particular day was (for any positive integer, i) between 20i and 20(i+1), 20(i+1) and 20(i+2), 20(i+2) and 20(i+3), 20(i+3 and 20(i+4), and between 20(i+4) and 20(i+5) units of flow measurement, respectively.

These latter two examples demonstrate that tallic numerals can provide extremely concise compilations of complete observational data in an error-checkable way. They leave much to be desired, however, with respect both to their pictorial charccteristics and to the ease with which particular quantitative values can be discerned manually. Evidentally these attempts to obtain a "double exposure" of both shades and shadows pictures and quantitative contours suffer too much from excessive mutual interference.

SIPLIC DATA

A more recent and evidently more successful trial[7] that eliminated these disadvantages resulted in the portrayal of hourly precipitation data reproduced in Figure 6. This particular form of tallic data is designated as being "siplic" data since it utilizes a "*s*caled *i*ncremental, *p*eriodically *l*abelled, *i*ncrementally *c*ontinuous" notation. It utilizes a relatively simple kind of tallic numerals to produce readability of quantitative values without destroying the pictorial "shades and shadows" character of incremental data.

As in the example in Figure 4 of the same precipitation gage data, 365 (or 366) lines, each consisting of 24 side-by-side tally marks, identify each hour of each day of each year. In this case, however, three transverse widths (instead of thicknesses) of tally marks are used to indicate (1) the occurrence of at least 0.01 inch of precipitation during the hour with a wide mark, (2) the occurrence of an as yet unaccounted for precipitation in excess of 0.01 inch in a closely preceding hour with a medium width of mark, or (3) no precipitation during the hour and no previously unaccounted precipitation with a narrow mark.

This incremental data is then "scaled" by inserting a medium-width scaling mark in the space behind those increments at which the accumulation reaches i(0.1) inches (for any integer, i), and a wide scaling mark in the space behind those increments at which the accumulation reaches i(1.0) inches. This technique of indicating values of higher order digits, rather than interfering with the "shades and shadows effect," actually enhances it. In regions of especially heavy rates of precipitation the several "extra" scaling marks tend to make the portrayal appear even darker, and the variations in shade of the portrayal are thus determined almost entirely by the rates and durations of precipitation.

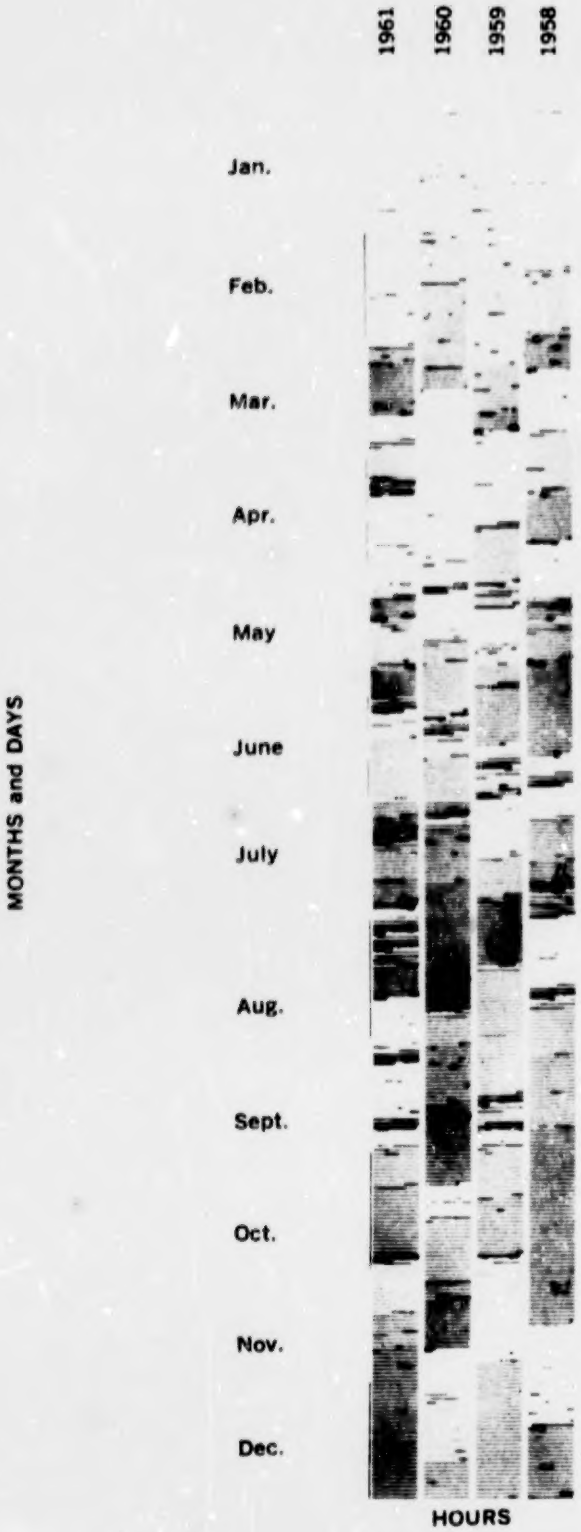
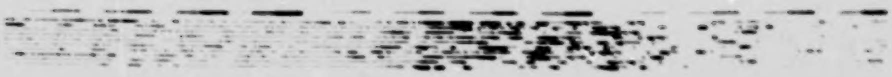
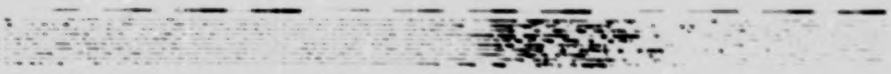


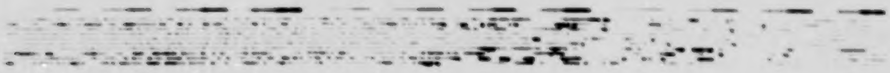
Figure 4 — Tallic Notation of Hourly Precipitation Amounts in Laramie, Wyoming, 1958 through 1961.



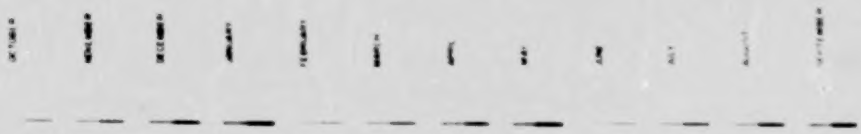
GREEN RIVER — UNIT OF RESOLUTION: 50 CUBIC FEET PER SECOND



ENCAMPMENT RIVER — UNIT OF RESOLUTION: 5 CUBIC FEET PER SECOND



SYBILLE CREEK — UNIT OF RESOLUTION: 0.5 CUBIC FEET PER SECOND



MONTHLY AND DAILY TIME SCALE

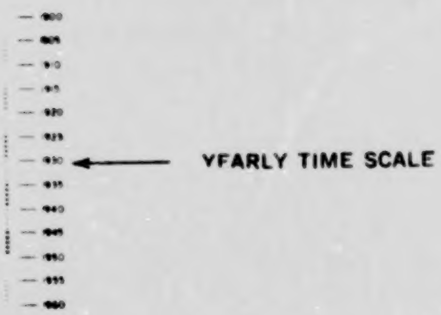


Figure 5 — Tallic Notation of Daily Values of Stream Flows.

In addition, "periodic labelling" of the values of higher order digits greatly enhances the ease of discerning quantitative values of accumulation throughout each year. It is accomplished with the pentiadic labelling lines alongside each year of incremental record. The zero, one, two, three, and four widths of numerals represent values between 5i+0 and 5i+1 inches, 5i+1 and 5i+2 inches, 5i+2 and 5i+3 inches, 5i+3 and 5i+4 inches, and between 5i+4 and 5i+5 inches of accumulation, respectively. The particular hour of the day in which any full inch of precipitation has accumulated is indicated directly by the position of a wide scaling mark in the incremental portion of the portrayal

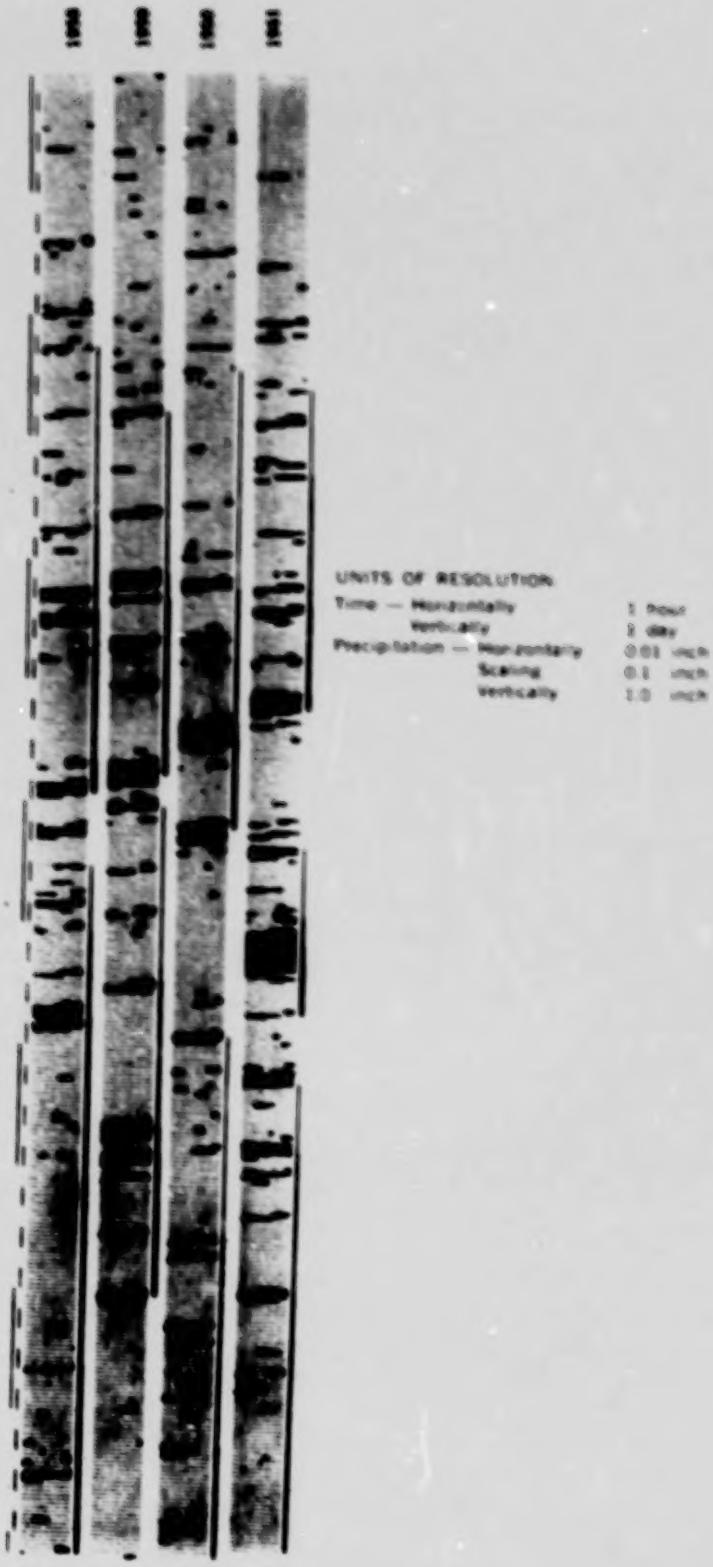


Figure 6 — Scaled Incremental, Periodically Labeled Notation of Hourly Precipitation Amounts in Laramie, Wyoming, 1958 through 1961.

for that day. The total accumulation for the year can easily be ascertained by first counting the numbers of 5's of inches corresponding to each major change from four-width to zero-width numerals.

CONCLUSION

It is concluded from these examples that the development of equipments appropriate for acquiring and utilizing siplic forms of tallic data should be very worthwhile. Evidently, they would be adaptable to portraying measurements of most if not all kinds of environmental conditions much more completely, concisely, and usefully than has been possible heretofore. Or, as previously discussed in more detail,[2] the development of such "informatic" kinds of equipments should make the ultimate goal of acquiring and utilizing "portrayals of everything geospheric everywhere always" much more approachable.

It is important in this respect to notice that although the incremental "pictures" in Figures 1 and 6 are at least as pictorial as many graphical "analyses" of contemporary meteorological and hydrological conditions they are in reality more nearly records of "raw measurements" than a result of "analysis." At least, they would be if appropriate observational instruments were available to acquire these kinds of continuous data at environmental measurement stations and if correspondingly appropriate equipments were available for collecting, compiling, and utilizing the observations in this form as outlined in Figure 7. Clearly, the key to opening this door to better understanding and utilization of our environment is the availability of a wide variety of continuously recording informatic observing instruments. Their development is now being emphasized in the University of Wyoming's program of Informatic Data Research.

REFERENCES

1. "Informatic Forms of Data, 1961," John C. Bellamy, Natural Resources Research Institute, University of Wyoming, November 1961, 11 pp.

2. "Geospheric Data Systems," John C. Bellamy, Natural Resources Research Institute, University of Wyoming, November 1961, 11 pp.

3. "Study of Usefulness of Unitary Differential Notation for Storing and Utilizing Meteorological Data," Cook Research Laboratories, Report No. 62-1, Contract No. AF 19(604)-1108, June 1955.

4. "Periods of Records of Stream Gages in Wyoming, 1890-1961," Philip M. Hoyt and John C. Bellamy, Natural Resources Research Institute, University of Wyoming, August 1962, 8 pp.

5. "Informatic Precipitation Gage Data," Merlin C. Williams and Leonard B. Baldwin, Jr., Natural Resources Research Institute, University of Wyoming, Septmeber 1962, 14 pp.

6. "Informatic Stream Gage Data," Verne E. Smith, Natural Resources Research Institute, University of Wyoming. (To be published).

7. "Siplic Form of Precipitation Gage Data," Anton C. Munari and Merlin C. Williams, Natural Resources Research Institute, University of Wyoming. (To be published).

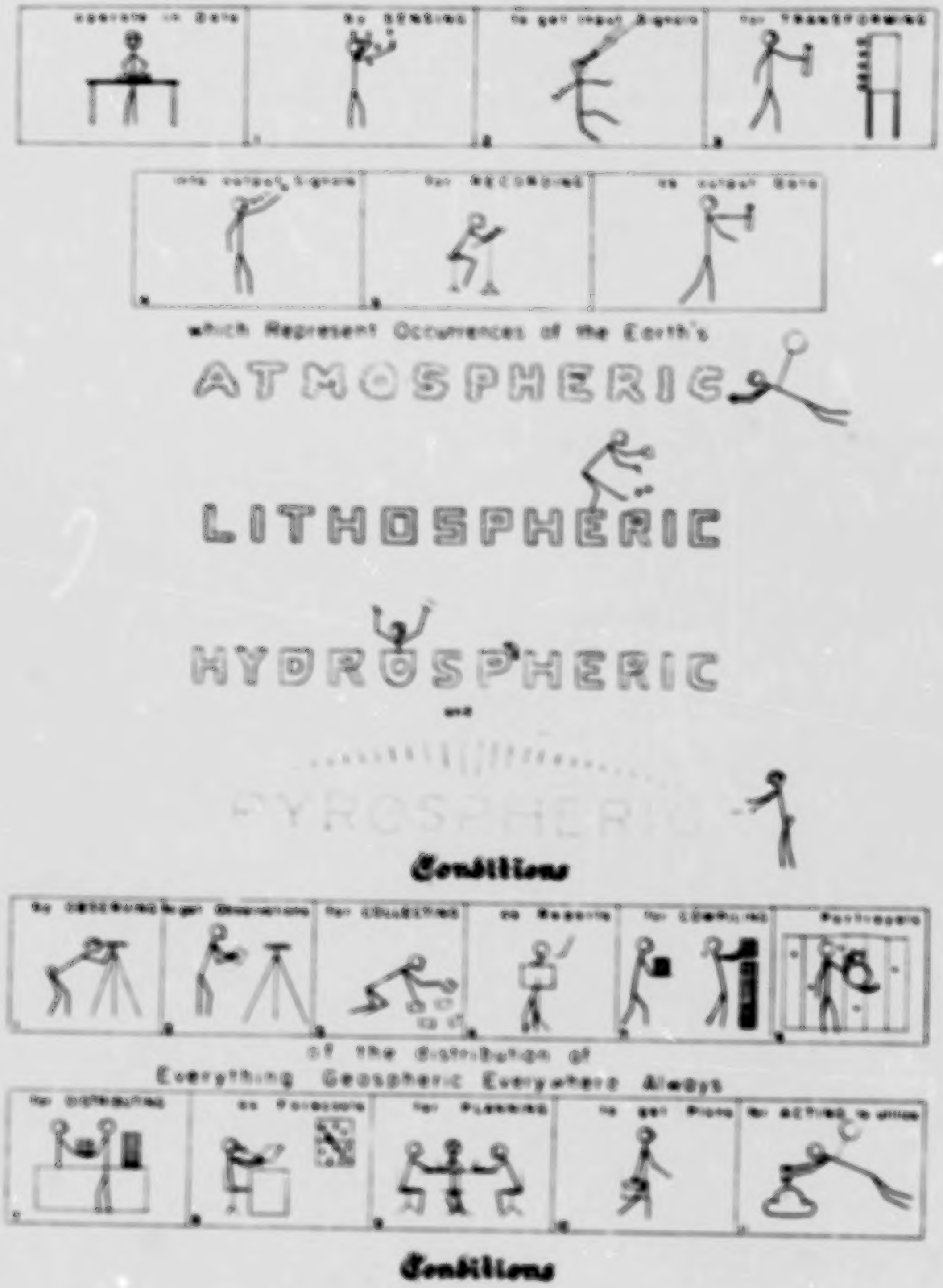

operate on Data
by SENSING
to get Input Signals
for TRANSFORMING
into Output Signals
for RECORDING
as Output Data
which Represent Occurrences of the Earth's
ATMOSPHERIC
LITHOSPHERIC
HYDROSPHERIC
and
PYROSPHERIC
Conditions
by OBSERVING
to get Observations
for COLLECTING
as Reports
for COMPILING
Portrayals
of the distribution of
Everything Geospheric Everywhere Always
for DISTRIBUTING
as Forecasts
for PLANNING
to get Plans
for ACTING to utilize
Conditions

Figure 7 — Geospheric Data Systems.

DISCUSSION

Dr. Larsen asked whether Dr. Bellamy knows of machines that produce good half-tone pictures of the data displays described. Dr. Bellamy replied that some are in the making at the University of Wyoming in connection with Masters and Doctoral work. Very little machinery on the market is adaptable directly. Some of the very complex and expensive machines completely invalidate the simplicity of this display technique.

Mr. Gelmont asked whether Dr. Bellamy finds difficulty in going from an analog signal to the actual printout without getting into some involved programming or computational procedures. Dr. Bellamy indicated that the main requirement is a good analog-to-digital converter and a little memory. The basic thing you are doing is keeping track of the previous value so you know what the step has been. This can be incorporated into the analog-to-digital mechanism by means of a stepping servo. Mr. Gelmont then asked why the data should not be stored on magnetic tape to provide a computer printout, which presents the topology of the situation. Dr. Bellamy pointed out the difficulty of identifying what is on a magnetic tape without running it through something. He stated that the density of storage in his system is compatible with or even better than most magnetic tapes. Potentially it can be played back at least 10 times as fast as magnetic tapes.

Mr. Linsky commented that a major advantage of this system is inexpensive reproduction in large quantities.

Glenn W. Brier
Chief, Meteorological Statistics Research Project
U. S. Weather Bureau, Washington, D.C.

SUMMARY

In the analysis of experimental data, the problem is to separate chance effects from true regularities. By the use of the probability theory, certain mathematical models are constructed that seem to bear at least some resemblance to the real world. This has led to many useful techniques such as Least Squares Curve Fitting, Analysis of Variance, Regression and Correlation Analysis, X^2 Goodness of Fit, etc. Examples are given that illustrate some of these techniques of data analysis; some aspects of extreme values are considered in the examples. Several morals are drawn from this discussion: a knowledge of the physics of the situation is necessary before a meaningful variable or parameter can be chosen for statistical analysis; there is value in knowing something about the observations — how they are taken, the peculiarities of the instrument or the observer, etc.; and, so that the real effects are not confused with statistical artifacts that could arise from data that are essentially random, the method to be used for processing the data should be understood so that a valid interpretation of the results can be made.

TECHNIQUES FOR DATA ANALYSIS

This is a very broad topic and time does not permit a thorough discussion of even a small fraction of the techniques available. Some of them are discussed in other papers at this symposium and numerous textbooks are available.[1–3,6] The emphasis here will be on general principles, and good texts along this line are also available. For example, "An Introduction to Scientific Research" by E. B. Wilson, Jr.,[7] is an attempt to explain simply a number of general principles, techniques, and guides for procedure.

Generally speaking, we are concerned with the analysis of experimental data. The problem is one of separating chance effects from true regularities and is treated as a branch of the theory of sampling. By the use of probability theory, certain mathematical models are constructed that seem to bear at least some resemblance to the real world. This leads to many useful techniques such as Least Squares Curve Fitting, Analysis of Variance, Regression and Correlation Analysis, X^2 Goodness of Fit, etc. It is doubtful whether a further enumeration of such statistical techniques or even a brief description of them is what we want here. A particular rule or a formula can be given, but there is no assurance that it will be applied correctly or chosen wisely. I think what we really want are "trained brains, and not a knowledge of facts and processes crammed into a wider range of untrained minds," as expressed by Karl Pearson. Or as Francis Bacon would have it, "minds . . . versatile enough to catch the resemblance of things (which is the chief point) and at the same time steady enough to fix and distinguish their subtler differences."

With these thoughts in mind, I have chosen some examples from the experience of myself and my colleagues to illustrate some techniques of data analysis that, we hope, can lead us to some general principles or conclusions. The printed program mentions "extreme values," so perhaps I won't be departing too far from the spirit or intent of the program to discuss some aspects of extreme values in these examples. The context in which I discuss extreme values is, however, very different from the one commonly encountered in statistical practice. The statistics of extreme values has become a specialized topic with at least one book[3] devoted to this topic alone. A typical application of the theory is directed toward the problem of estimating the probability that some natural phenomenon such as an extreme flood will occur within a specified period of time or that a piece of machinery will fail or break down. Also, studies have been made about rules relating to the rejection of observations of extreme values that do not appear to "fit in" with the rest of the sample.

The first example here refers to a series of measurements of solar intensity by means of an instrument called the pyrheliometer. This instrument measures the intensity of the direct solar beam at the surface of the earth and therefore is affected by the atmosphere that contains dust and cloud particles, smoke, water vapor, etc. Instructions to the observers at U. S. Weather Bureau stations are that observations should be taken at specified solar zenith distances only when there are no clouds obscuring the sun. This is somewhat subjective, since one observer may take more observations in a period of time (such as 1 month) than another observer. When monthly averages are taken, they will tend to run higher for the observer with fewer observations, since he has selected only the "clearest" skies. Although both observers have chosen the extreme values, in a sense, one of them has included observations closer to the mean or median of the frequency distribution. One effect of this is shown by the graph for Blue Hill in Figure 1. A new observer came on duty near the beginning of 1952. He has chosen only the "clearest" skies, so that the "average" appropriate to his observations is approximately 11 percent higher than the "normal" for the period 1934 through 1951, shown as the heavy "O" line. This bias might be avoided by using only the highest value each month, and this might be very desirable if long-term trends in the data were being studied and more than one observer was involved. Figure 1 came from investigation of the question whether a volcanic eruption in Alaska in July 1953 produced extensive pollution in widespread regions of the earth's atmosphere.[4] The data for both Blue Hill and Table Mountain shown in Figure 1 give some support to the suggestion that such an influence existed during the last part of 1953 and the first part of 1954.

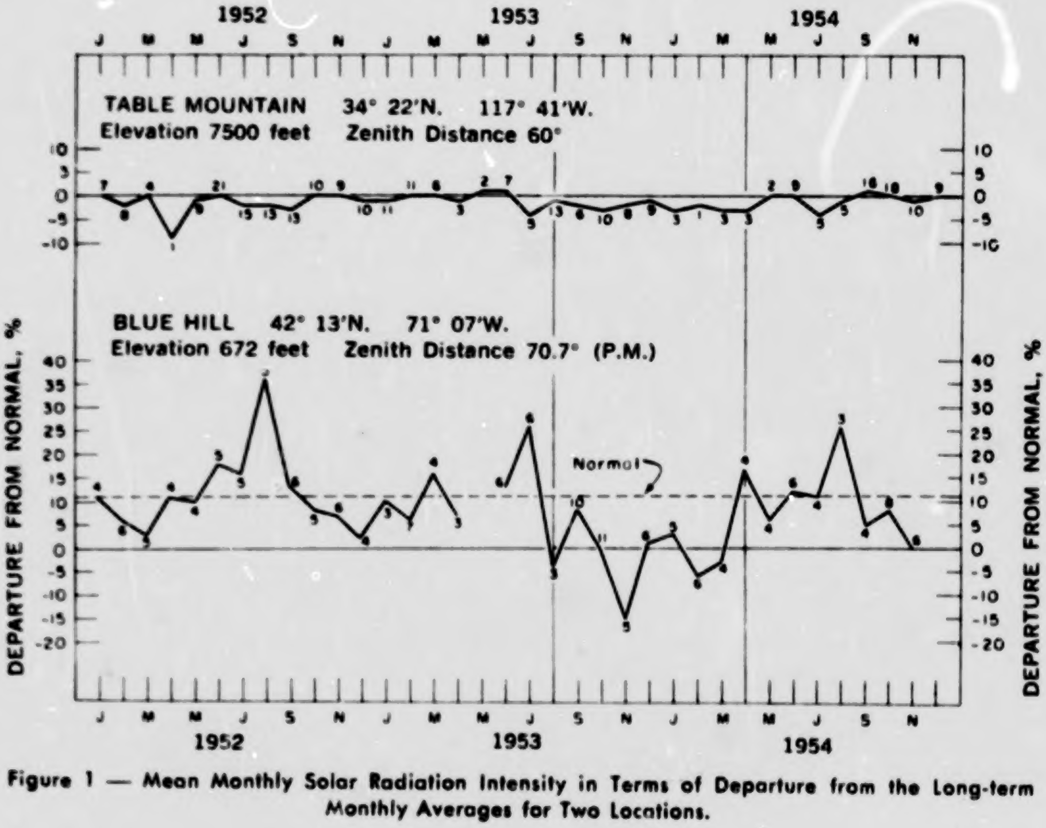


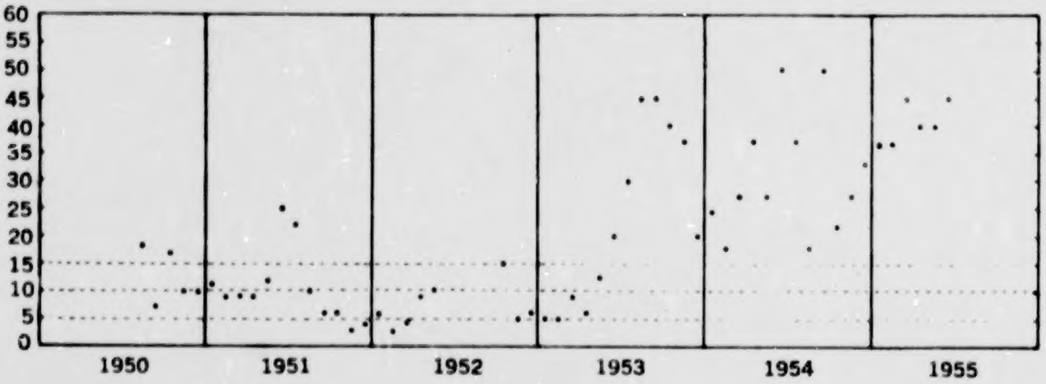
Figure 1 — Mean Monthly Solar Radiation Intensity in Terms of Departure from the Long-term Monthly Averages for Two Locations.

It was possible to investigate this same question by a different type of data. Sky photometer readings were available from Climax, Colorado, and Sacramento Peak, New Mexico, for the period 1950 through 1955. This instrument measures the intensity of the light from the sky at angles near the sun and compares this intensity with the intensity of the direct solar beam. Dust or other scattering particles in the atmosphere tend to increase the sky readings, whereas, very low readings indicate the absence of particles due to dust, smoke, clouds, etc. Table 1 shows a sample of the original data used.

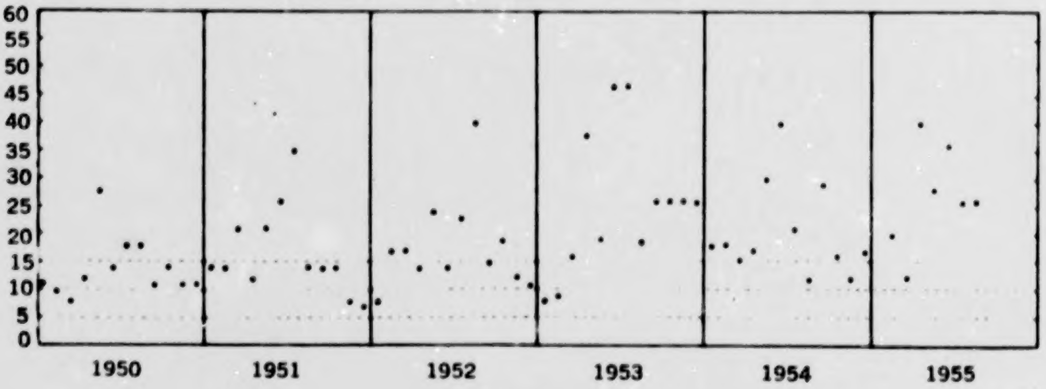
Table 1 — Sky Photometer Readings at Climax, Colorado, and Sacramento Peak, New Mexico, for January 1953.

Day	Climax		Sacramento Peak	
	Greenwich Time	Sky readings	Greenwich Time	Sky readings
1			1718	62
2			1655	26
			1759	35
			1809	35
			1844	23
			1902	23
3			1550	35
			1607	35
			1644	23
			1702	23
4			1515	90
			1531	90
			1729	16
			1745	16
5			2105	42
			2120	42
8			1740	15
			1758	15
9	1600	5	1524	35
	1651	5	1542	31
	1717	5	1545	31
	1745	5	1654	21
	1833	5	1705	31
	2019	5	1712	21
	2048	5		
	2159	15		
10			1510	38
			1533	38
			1542	38
			1705	14
			1719	13
			2129	26
11			1540	31
			1633	14
			1650	14
			1853	35
12			1630	29
			1645	30
			1716	>500
13			1530	52
			1550	50
			1555	50
			1926	29

Considerable variability is indicated, and the occurrence of values of 500 or greater would have a large influence on a daily or monthly mean. For the purpose of this study, it was reasoned that selecting the lowest value each month would make considerable physical sense. If the eruption of the volcano in 1953 produced an extensive pollution of the upper atmosphere over widespread areas, then extremely low values of sky brightness should no longer be observed because of the ever present common background of extra particles producing scattering in the atmosphere. Figure 2 suggests that this is actually what happened. There is a seasonal factor, with a deficiency of low values during the summer months, but the winter of 1953-1954 shows an absence of low values for both Climax and Sacramento Peak. The Sacramento Peak data suggest a return toward normal seasonal conditions by the end of 1954, whereas the Climax data for 1954 and 1955 suggest the possibility of a "drift" in the instrument. On the basis of the results of these charts, technicians examined the Climax photometer and found that it needed to be recalibrated!



CLIMAX



SACRAMENTO PEAK

Figure 2 — Lowest Values of Sky Photometer Recordings Each Month for Climax and Sacramento Peak.

Although these studies show the value of using only the most extreme observation taken during an interval of time, the objection might be made that the extreme value may depend too much on an instrumental or observational error. For example, if errors due to mis-reading the scale of the instrument or transcribing the data are frequent and large, the extreme values may be extreme only because of "goofs." The researcher must know something about the magnitude and relative frequency of such errors before he can make rational decisions involving the treatment of extremes. In some cases, it might be better to ignore the most extreme value and take the second highest (or lowest), for example. The optimum procedure will depend upon the physical and other factors involved. Figure 3 shows for Sacramento Peak the lowest daily readings of the sky photometer for January 1953 and December 1953. In this case, it would not make much difference what statistic is used to compare the 2 months. January would be lower than December whether the highest or lowest values were chosen, the second highest or second lowest, etc.

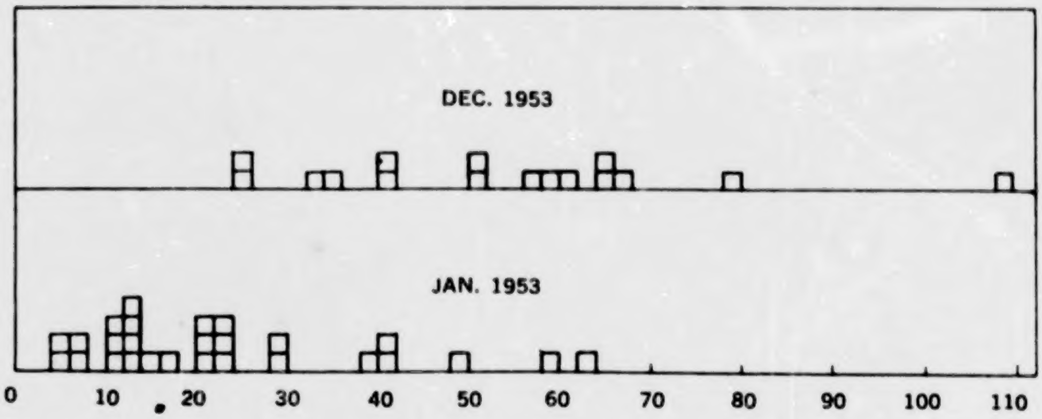


Figure 3 — Lowest Daily Readings of Sky Photometer for Sacramento Peak During January 1953 and December 1953.

In another type of investigation, it is common to use extremes for correlation or regression studies. This can have important effects on the interpretation of the results, especially when it is not recognized that extremes are being used or the investigator is not aware of some of the subtleties or pitfalls in using extreme or rare events. For example, in weather forecasting, we might wish to study how the frequency of extremely large daily rainfalls is related to surface dewpoint. For a medical or health study, we might wish to study how the frequency of patients with blood pressure over 200 is related to the amount of salt in the diet, for example. The relationship between a dependent variable of and a possible causal or independent variable x can be represented by a scatter diagram like that shown in Figure 4. In this figure, all the data are shown for x and y, not just the extremes. If only the highest values of y are considered, however, such as those above the line $y = k$, it can be seen that the relative frequency of these events can change very rapidly as x changes from negative values to positive values. If we assume that x and y are distributed in a normal bivariate distribution with means $\bar{x} = \bar{y} = 0$ and standard deviations $\sigma_x = \sigma_y = 1$, it is easy to construct Table 2 to show quantitatively how the relative frequency of an extreme event y changes or x varies. This depends upon the average frequency ($\overline{py}$) of the event y and upon the correlation r between x and y. Even with a correlation as low as $r = 0.01$, there is considerable change in the relative frequency of an event according to the value of x. For example, if we take the case where an event happens only 13 times out of 10,000 ($y = 35$, $\overline{py} = 0.00135$), we find that it is $100 \times \frac{(118-103)}{103} = 14.6$ percent more likely, when x is $2\ \sigma$ above the mean than when x is $2\ \sigma$ below the mean. For a correlation as high as $r = 0.10$, the percentage variations are tremendous, even though a correlation of $r = 0.10$ means that 1 percent of the variance of y is accounted for by the regression of y on x. Thus, extreme caution must be used in interpreting the results of investigations where the frequency of unusual or extreme events (the pathological

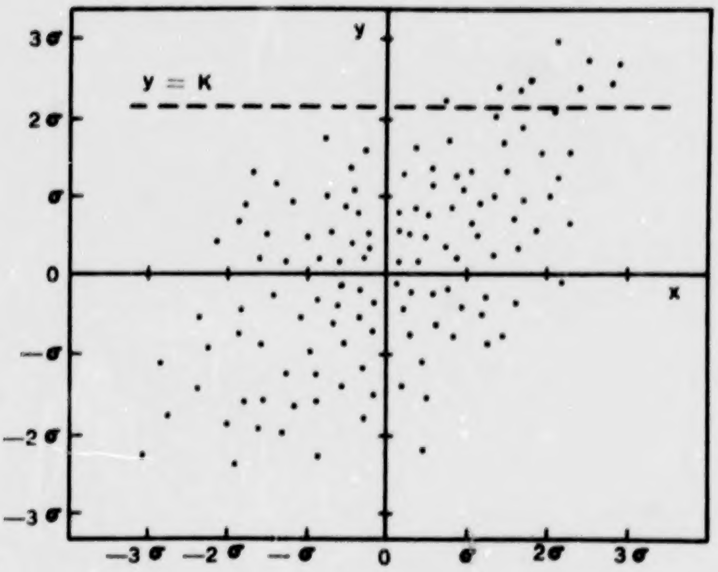


Figure 4 — Scatter Diagram Illustrating the Relationship Between Variables x and y.

conditions) are related to some asserted causal variable. The use of extremes appears to sharpen up the relationship or amplify an effect. It is important that the investigator know what he is doing and understand the consequences that result from a particular treatment of the data.

Table 3 — Change in Relative Frequency of an Extreme Value of y as a Function of x for a Normal Bivariate Distribution.

	r = 0.01			r = 0.10
x	$y \pm 1.83$ $(\bar{P}y = 0.033333)$	$y \pm 3$ $(\bar{P}y = 0.001350)$	$y \pm 4$ $(\bar{P}y = 0.000032)$	$y \pm 3$ $(\bar{P}y = 0.001350)$
−3	100	100	100	100
−2	102	103	104	141
−1	105	107	109	200
0	107	110	113	278
1	109	114	119	374
2	112	118	123	527
3	114	122	129	714

Now what are several morals we might draw from the preceding discussion. I think there are at least three. First, we see the importance or the necessity of a knowledge of the physics of the situation before a meaningful variable or parameter can be chosen for statistical analysis. In the first examples given here, it made considerable sense to use only the extreme values, whereas in other circumstances these might be discarded entirely. Second, we see the value of knowing something about the observations — how they were taken, the peculiarities of the instrument or of the observer, etc. Third, we see (I hope) that valid interpretation of the results cannot be made without an understanding of the method of processing the data. That is, we must know the consequences of our particular way of treating the data so that we will not confuse real effects with statistical artifacts that could arise from data that are essentially random.

REFERENCES

1. Bowker, A. H., and Lieberman, G. J., "Engineering Statistics." Prentice-Hall, Inc. (1959)
2. Brooks, C. E. P., and N. Carruthers, Handbook of Statistical Methods in Meteorology. New York, British Information Services, 845 Third Ave., (1953).
3. Fisher, R. A., "Statistical Methods for Research Workers." Oliver and Boyd, Edinburgh (1941).
4. Fritz, S. "Opacity of the Atmosphere after July 1953," Meteorological Magazine, vol. 85, No. 1006, April 1956, pp. 311-312.
5. Gumbel, E. J. "Statistics of Extremes." Columbia University Press, New York (1958).
6. Panofsky, Hans, and Glenn W. Brier, Some Applications of Statistics to Meteorology. University Park, Pa., Penna. State University, (1958).
7. Wilson, E. B. Jr. "An Introduction to Scientific Research." McGraw-Hill, New York (1952).

Dr. Donald W. Pritchard
Director, Chesapeake Bay Institute
Johns Hopkins University, Baltimore

SUMMARY

The dangers of manipulating data to conform to an established hypothesis are illustrated, in particular the use of data both to formulate hypotheses and to verify them. Data should be collected to provide answers to clearly defined, specific questions that we ask about the environment, questions that will prove or disprove a given hypothesis. Methods of data collection must be compatible with the techniques of interpretation that we intend to apply.

INTERPRETATIONS AND CONCLUSIONS

A reading of the program of this symposium indicates that others are scheduled to speak specifically about data interpretation and conclusions for the separate environments of air and water. I am here apparently scheduled to present some "basic" statements applicable to the general subject of interpretation and conclusions. I have interpreted this situation as giving me the license to speak rather broadly on the philosophy of interpretation of environmental data.

In what follows, I am indebted to my colleague Dr. Blair Kinsman, who has written very eloquently on this general subject in his paper "Proper and Improper Use of Statistics in Geophysics" (Kinsman, 1957). The first portion of my presentation here is essentially a paraphrase of a part of Dr. Kinsman's paper, since his thoughts on this subject so nearly coincide with my own. Where I have found that no rephrasing on my part adds to the clarity (to me), I have simply, and perhaps lazily, quoted directly from Kinsman.

The environmental scientist shares with all other scientists the task of telling "likely stories" with the intent that the story as told will satisfy the observations that the scientist has made. The concept that this is a business of scientists is an old one, dating back at least to Plato. The materials, that is, the set of data or observations, with which a scientist works are "appearances", that is, sense impressions, and, as expressed by the old Greek phrase, he tries to "save the appearances." Given a finite set of observations, this is a fairly straightforward task. "An excellent example is offered by Ptolemy, who takes astronomical observations back to the reign of Nabonasser and shows that he can 'save' them, that is fit them into a coherent pattern, by telling either a geocentric story or a heliocentric story. With two equally satisfactory stories Ptolemy weighs their likelihood and, on the basis of the physical information available to him, develops the geocentric story as the more likely. The basis for this choice is hardly so simple and straightforward. Today the general consensus, except possibly among navigators, is that the heliocentric story is the 'true' one.

"The advent of the word 'true' together with the word 'real' in scientific discussion has done much to cloud the nature of scientific activity. The request for a 'true' story instead of for a 'likely' story tacitly postulates the existence of a 'real' world underlying and giving rise to appearances and asks for information about that 'real' world. The scientist *qua* scientist cannot answer such a question since the material on which he operates consists entirely of appearances. St. Thomas Aquinas would probably have said that no one could answer, since 'Nothing was ever in the mind that was not first begot upon the senses.' The attempts to bridge the gap between the postulated 'real' world and the world of appearances which we perceive has a long and uniformily unsatisfactory history, covering the spectrum from Descartes' assertion that God would not fool us to Berkeley's retreat into solipsism. If we restrict ourselves to 'appearances,' 'hypotheses,' and 'likelihood,' metaphysical speculation about 'truth' and 'reality' can be left to the metaphysician with a considerable gain in clarity.

"The point of departure is the appearances. These range from the casual impressions of any sentient being through the systematic observation of essentially uncontrollable phenomena characteristic of . . . ," for example, environmental science, "to the precise measurement of the results of highly controlled experiments characteristic of the laboratory sciences. The habit of attentive observation, coupled with an overwhelming urge to fit the observations to a pattern, embryonic in most of the human race, is developed in the scientist to a high degree. All appearances, however, are not suitable for scientific activity. Aristotle said that the subject matter of science is that which happens always or for the most part. The unique event is a subject for history. Poincare (1905) puts it this way:

> Carlyle has written something after this fashion. 'Nothing but facts are of importance. John Lackland passed by here. Here is something that is admirable. Here is a reality for which I would give all the theories in the world.' . . . The physicist would most likely have said: 'John Lackland passed by here. It is all the same to me, for he will not pass this way again.'

"Having, then, a set of observations of a recurring phenomenon the next step is to construct an intelligible hypothesis into which the observations can be made to fit. One fertile method is the use of analogy. Some other set of phenomena and their pattern being known, if we can see a similarity, we can transfer the properties of the known system to the unknown. Since analogies are seldom isomorphisms, the correspondences being only partial, the dangers of argument by analogy are obvious. For example, the complex numbers are analogous to the real numbers in the sense that operations of addition, subtraction, multiplication, and division can be defined for each. We might then argue by analogy that, since division by zero is not permitted for the reals, division by zero is not permitted for the complex numbers. We thus reach a correct result. If we argue by analogy that since the reals are ordered the complex numbers must also be ordered, our conclusion is false. Fertile as the argument by analogy is as a source of ideas, it is almost worthless in support of an hypothesis unless it is shown that the analogous systems are similar in every essential feature and that no dissimilar features can affect the properties that we wish to establish. Another method is to search the data for regularities. If the sample is small . . .", as is generally the case in the environmental sciences, "this is quite easy to do. However, since even samples drawn from a random number table will exhibit regularities, results from such a procedure are suspect and once a regularity is found most scientists feel impelled to provide some rationalization for it, often based on an *ad hoc* selection of arguments. . . . Perhaps the most acceptable method of forming hypotheses is by rational argument from established elementary principles. It is worth noting that scientists in general seem to feel uneasy about any hypothesis until it has been presented in this form no matter how it was first conceived. An argument in Kepler's 'Epitome' is a case in point. Kepler's hypothesis that the planetary distances are governed by the porportions of the five regular or Platonic solids seems a little wayward today, but the urge to order the welter of appearances is easily understandable. Kepler apparently felt the need to deduce his hypothesis from the first principles of geometry and astronomy since he devotes considerable space to the effort.

His logic is impeccable except for one short section. There the line of the argument has been blurred, whether consciously or unconsciously cannot be known, so that the ostensible deduction he was at such pains to make is, in fact, not established. The expenditure of so much effort in such a cause by a man with first-rate discoveries to his credit shows the importance attached to this method.

"But the telling of tales is only half the job and the easiest half at that. We still face the problem of deciding how likely the story is or, if confronted with two different but adequate stories, which is the more likely of the two. Statistics has been increasingly concerned with understanding the structure of such decisions and with finding a clear and objective method of making them. The general problem is far from solved, but many valuable results have already been obtained. The judgments of the likelihood of an hypothesis have had so many different bases that even a simple enumeration would be too long. They include decisions made on entirely extraneous grounds, e.g. the selection of a flat earth over a round earth on arguments derived from the 'second coming.' They include decisions made on what I should call aesthetic grounds, e.g. the selection of uniform circular motion as basic by the Greeks in contrast to uniform rectilinear motion by Newton. They include decisions in which maximum simplicity is equated with maximum likelihood. Occam's Razor is still a widely used scientific tool, although the simplicity of nature is more an article of faith than a proven fact. The more we refine our studies of nature the more complex things become until it almost seems as though simplicity were an attribute of the infancy of a science. Fortunately, there are a few threads to guide us in the labyrinth. Occam's Razor has been mentioned. If you are willing to commit yourself to the proposition that the relations among phenomena are fundamentally simple, then you will always choose the simplest story that explains all the facts. However, it must always be borne in mind that one man's simplicity may be another man's utter confusion. A heliocentric hypothesis simplifies the astronomer's calculations but it makes those of the navigator intolerably complex. If one is willing to forego questions of 'truth' and 'reality' one can escape the dilemma by accepting both the geocentric and the heliocentric hypotheses (so long as they are not in logical conflict) and use whichever one is simpler *for the immediate purpose*. A most important criterion is that of compatibility with already existing structures. If the hypothesis under consideration would require extensive revision of major parts of a successful existing theory with all the labor that entails, clearly one would hesitate to accept it unless a general improvement throughout the whole theory could be anticipated. Another equally important basis for decision is the continued agreement of observations with the hypothesis. Here statistics enters, since no set of experimental measures, if sufficiently refined, ever agrees exactly with an hypothesis or with other sets. The tincture of statistics that most of us retain from our formal training seldom goes beyond the memory of where to find a few computing formulae. This isn't enough. While a sprinkling of Pearson's χ^2's, Student's t-tests, and Fisher F-tests do lend an air of objectivity to any paper, it must be remembered that statistical tests of significance derive from mathematical models, which in turn are based on different views of the nature of phenomena. The calculation of parameters is routine, and their use often obscures a lack or precise thought about the fundamentals of a problem.

"In general, the job of the scientist is to invent a story which accounts for a set of observations and then to decide how likely the story is.

"The . . ." environmental "sciences share the same theoretical base with laboratory sciences. The Navier-Stokes equations hold equally well for a beaker of water and for . . ." a river, a lake, or "the oceans. With this common base it is not surprising that the methods for treating . . ." environmental "data *are often selected* by analogy with those used on laboratory data. Unfortunately, the materials on which these methods are used in . . ." environmental science "are sufficiently different from those of the laboratory sciences to require justification of the method, which is seldom explicitly given. In the first place, the equations which give a full description of any situation are usually too complex to be handled in their complete form. It is almost always necessary to simplify them by considering some terms as negligible in order to get an approximate solution. The laboratory scientist, by controlling the conditions of an experiment, can insure that terms considered negligible are so in fact and, as a result, he can expect good agreement between hypothesis and experiment even with relatively small samples. Further, he can repeat his experiment at will.

"The . . ." environmental scientist "who, in the main can only observe, cannot repeat his observations in the sense of a repeated experiment. This is a grave difficulty, since most statistical tests of significance are fundamentally rooted in the idea of repetition. While such repetition is conceptually possible in . . ." environmental science "it is seldom realized and, in general, when such tests are used, their use should be supported by argument. Further, his inability to insure that factors considered negligible are so should logically force the . . ." environmental scientist "to use the equations for a process in their complete form or first to show that the neglected factors can be neglected. Neither of these courses is usually taken. There is nothing wrong with the making of simplified models so long as they are not offered as 'reality.' When the observations happen to agree with such models it is cause for gratification and suggests that the neglected terms were small. If the observations continue to agree with the model we can feel that we have a satisfactory story. In . . ." environmental studies "the agreement usually need not be very close before verification is claimed. . . . This attitude stems from lack of control. In comparing data with a simplified model large dispersions are to be expected. This means that large samples are necessary if relations are to be established with any certainty. It is unfortunate when the investigator, having been forced to accept a simplified model, then feels impelled to insist that the natural phenomena are themselves simple.

"Another difference arises from the answer that can be given to the question: 'Do the *data* describe the phenomenon under consideration?' In the laboratory science methods can usually be devised either to measure a property directly or to measure some closely linked property. With the method in hand, the experimenter can then accumulate enough data for a statistically reliable estimate of the property which interests him. The . . ." environmental scientist "here labors under two kinds of handicap. First is the matter of scale. Both the space and time scales are usually unwieldy. If one asks for the monthly average temperature of the Chesapeake Bay is it enough to dip a thermometer in once a day at some convenient place? One can hardly say without knowing a great deal about the structure of the Bay. If it were enough for, say, January 1969, would anything be known about January 1970? To answer such questions an inordinately expensive observation network would have to be established and maintained for many years. Salinity records taken daily at Solomon's Island, Maryland, are a case in point. This set of data extending back to 1938 is the longest unbroken record of salinity taken anywhere on the Bay. Using monthly means and computing the power spectrum it was found that there was evidence of a yearly cycle, which was to be expected. However, the great bulk of the power in the signal occurred at periods greater than two years. It was calculated that to separate cycles having periods of three years and four years at the 5% level, the record would have to extend over 265 years. Such a sample from . . ." an environmental science "point of view is huge and it is rather sobering to see how little information it gives. Second, the . . ." environmental scientist "must frequently work with data which were taken for other purposes and which do not directly measure the properties that interest him. For example, an oceanographer interested in the factors influencing the size of fish populations might not have any measurements directly made for that purpose but instead measurements of salinity made at some point in the region and records of commercial fish catches. It isn't what he wants. It's what he's stuck with. If he persists, he would have to argue something like this. Fish population controls fish catch. Fish need plankton for food. Plankton need dissolved nutrients. Nutrients are brought to the surface layer by upwelling. Upwelling influences salinity. The salinity of the region can be determined from the salinities measured at a point which I know. Therefore, I will look over fish catches and salinities for possible correspondences and, if I can construct one, I will know the connection between fish population and environment. Laying aside the questions of whether the salinity measured at a point represents the salinity over a large area and of whether fish catch is an adequate measure of fish population, it seems unlikely that definite clear-cut relation between the ends of such a long and tenuous chain would emerge from a small sample.

"Another hazard inherent in using existing data taken for other purposes arises from the temptation to fill gaps in it. If properties *A* and *B* are to be related and it is found that *A* was measured at some point for a number of years but that the measurement of *B* was neglected for a part of the time, then the urge to use measurements of *B* made somewhere else to fill the gap may be almost irresistible. This procedure enlarges the sample with an apparent increase in statistical reliability but it introduces tacitly the very difficult additional problem of showing that the measures introduced to fill the gap are the equivalent of what would have been secured had *B* been measured at the point. This is usually impossible. In using existing data for purposes for which they were not taken, great care must be exercised to see that wishful thinking does not govern the make-up of the sample."

Environmental studies generally need large samples. Usually only small samples are available. "In contrast with the laboratory sciences these small samples are often imprecise, having been painfully secured in the field over many years, sometimes two or three generations. To get another sample for testing involves the same long process. Thus, if the entire initial sample is used in the formulation of an hypothesis, we are forced to leave its verification or rejection to our grandsons. It is clear that progress of a science which must either proceed on untested hypotheses or wait for generations to test them will be either insecure or very slow.

"Verification is to be had only from data not used in formulating an hypothesis. One possible method of securing data from testing an hypothesis formulated by search, without delay, is to split the data on hand into two groups, one to be used in formulating the hypothesis and the other reserved for testing. This may be done in a number of ways. In some fields dealing with time series every other time unit is grouped to form the two sets, or the data may simply be split in the middle. Separation by means of some randomizing device could be used so that the bias, conscious or unconscious, of the investigator would not invalidate such statistical tests of significance as might be appropriate. The real difficulty here is 'keeping the game honest.' If the hypothesis is formed before the test data are taken no question of influence arises. With both sets of data in existence at the beginning of an investigation there is always the question of the extent to which the investigator is influenced in his selection of hypotheses by the test data. A glimpse of it, however fleeting, could bias him toward hypotheses likely to fit both sets. The difficulty could be met if the separation were made before the investigator saw the data and he inspected only one set until he was ready to test his hypothesis. Any alterations in the hypothesis after testing would, of course, be highly suspect. The advantage of this device is that verification can be carried out at once and an estimate of the value of the hypothesis made. The disadvantage is that the already small sample size is further reduced, but it may be worth accepting this reduction in exchange for immediate evaluation. It is well to remember that the information contained by any finite sample is limited. Manipulating it in this way cannot increase the amount of information contained. It can only sacrifice information of one kind to gain information of another."

One basic difficulty in treating environmental data arises from the fact that we seldom have two unique sets of values of specified parameters that may be paired in a clearly objective manner for comparison. Normally we have one finite set of observations, the variation in which we wish to "explain" in terms of the variations in a "controlling" environmental parameter. Observations of this "controlling" parameter make up a second set, which, after suitable manipulation provide a series of numbers that are paired with and compared to the first set. Putting aside the questions associated with generally circuitous and usually unprovable story we invent to show that the particular parameter chosen is actually a "controlling" parameter, we are faced with the fact that frequently all the statistical significance of the final results is destroyed by our use of the data themselves to determine what manipulations are suitable.

A simple example will illustrate this situation. Suppose we have a set of observations of the annual harvest of young "seed" oysters from a productive oyster bed for a continuous period of, say, 20 years. As is the nature of such data, we will find considerable year-to-year variation in the harvest. As is usually done, we now equate the annual harvest to the actual production and survival of the seed oysters on the bed in question. We now want to explain the year-to-year variations. Coincidentally we find that there have been, over this same 20-year period, daily observations of the salt concentration at the condenser cooling intake of an industrial plant located not too far (?) from the oyster bed. We conclude that daily observations are too variable and anyway provide too many numbers to work with, so we compute the monthly mean salinity of the environment at a location near our oyster bar.

We now have a series of 20 numbers representing the annual harvest of seed oysters, and a series of 12 x 20 numbers representing the monthly average salinity. It takes only a moderate amount of imagination, which we usually in such cases call reasoning, to invent a story that convinces us that the mean monthly salinity should "control" the production and survival of the seed oysters. In fact, we would probably reason that the salinity during one part of the year would influence the condition of the brood stock, and hence the number of eggs and sperm produced; while the salinity during another part of the year would influence the fraction of young oysters that survived to the time of harvest. Unfortunately our story is usually not complete. We are not sure which of the 12 monthly values of salinity in each year is most important from the standpoint of production of larvae and which is most important from the standpoint of survival. We therefore proceed to compare the observed oyster harvest to a computed harvest for each year based on a multiple regression of all combinations of pairs of monthly mean salinities from the 12 months just previous to the harvest. Hurrah! We find that if we use the monthly mean salinities for, say, the previous May and for December in our regression relationship the computed oyster harvest is highly correlated (on the order of 0.95) with the observed harvest for the 20 years of record. We have now "explained" the year-to-year variation in oyster harvest in terms of variations in an environmental parameter!

Unfortunately we have used the observed data to search for the best relationship. In point of fact almost any set of numbers showing some type of cyclic variation, such as the mean monthly value of some environmental parameter, can be made, through suitable manipulation, to show a high correlation to the annual variation in some other set of observations. The "explanation" of the variations in oyster production arrived at in the previous paragraph has, in fact, no statistical validity!

The number of published papers in which essentially the approach described above has been used to explain the variation of some property of the environment is considerable. Kinsman, in the work cited previously, analyzed a paper in which the author attempted to show that the number of icebergs counted by the Ice Patrol in the North Atlantic in any given year was related to the monthly mean sea-surface temperature anomalously obtained from measurements at the end of a pier at Key West, Florida. In order to show the relatively high probability of obtaining apparently significant correlations between finite series when in fact any physical connection is nonsense, as long as some choice for manipulation of one set is allowed, Kinsman counted the number of commas per page in the issue of the journal in which the original paper on icebergs was published. Kinsman correlated the number of icebergs in a given year to the number of commas per page in the subject journal, but left himself the option of proceeding either forward or backward in the page count, and of selecting which page he would start his comparison with. He found that when he computed the number of icebergs per year based on the number of commas per page in the journal, starting with the last page of the article he was analyzing and proceeding in page sequence toward the front of the paper, he obtained a correlation of 0.95 with the observed iceberg count for the years 1942 through 1951. The comparison is shown graphically in Figure 1, taken from Kinsman's paper. He then proceeded to use the relationship thus obtained, together with the number of commas per page, running backwards, in the article just preceding the one he had analyzed, to "predict" the iceberg count for succeeding years. As shown in Figure 1, the prediction for the three years 1952, 1953, and 1954 is quite good. Thereafter, as would be expected, the prediction failed completely.

Evidently Kinsman's selection of data was fortuitous; however, this example does serve as a vivid warning about the way environmental data are often used.

It is my experience that most environmental data have been collected under programs developed without adequate consideration of how the results will be used. The time has come when we should severely limit the amount of effort being expended on the general collection of environmental data, for which we have only vaguely or partially conceived the use.

What, then, should our course of action be? First we must recognize that from a practical standpoint it is impossible to develop a single over-all observational program, involving even a limited number of environmental parameters and a restricted natural environment, that will provide data suitable for use in answering all, or even a considerable fraction, of the questions that need to be answered regarding the subject environment. The methods and timing of data collection suitable for the treatment of one question about the environment will seldom be satisfactory for dealing with other questions. Data collection programs designed without regard to specific, completely stated questions that we want to ask of the environment will generally be not quite adequate to definitely answer any question.

Basically, then, the subject of interpretations and conclusions, which appears near the end of this symposium program, should in fact be an integral part of the initial development of an environmental observational program. The first step is to state clearly

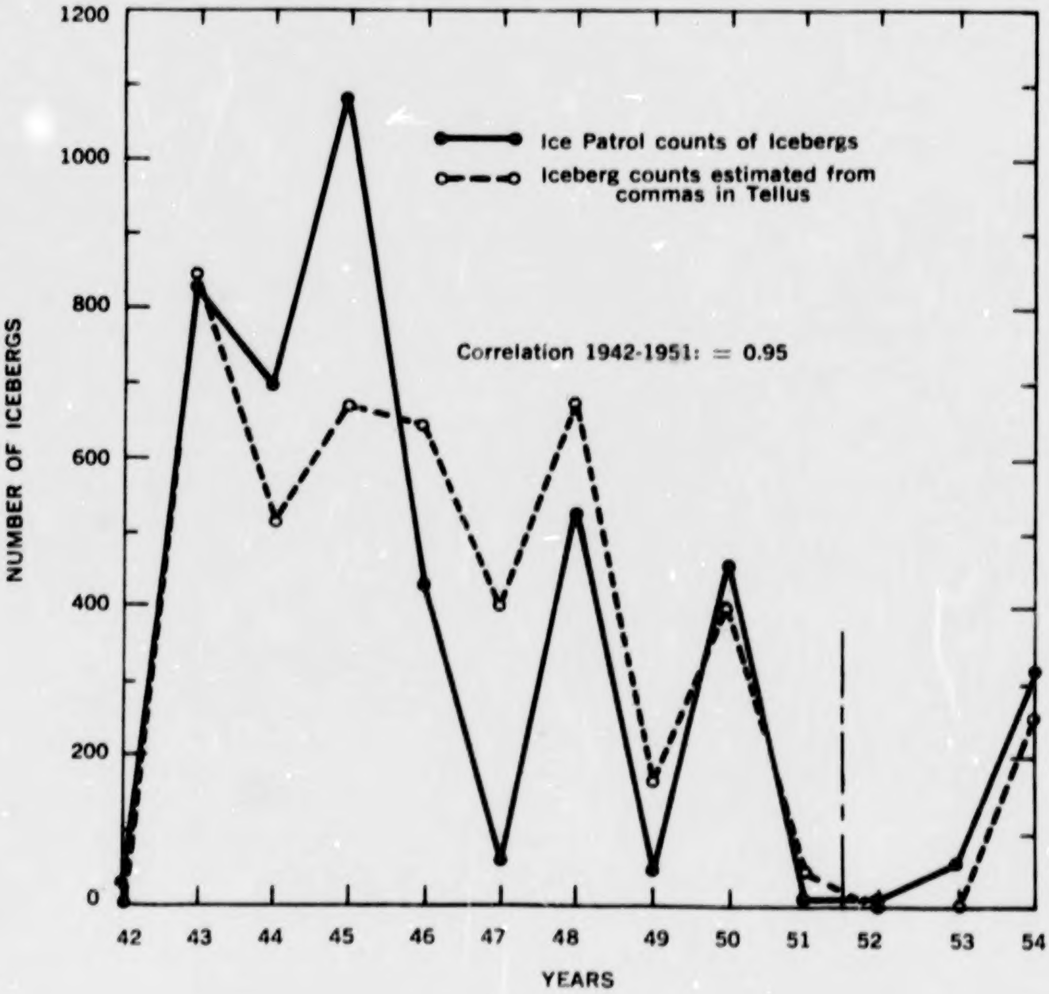


Figure 1 — Correlation of Number of Icebergs in a Given Year to Iceberg Counts Estimated from Commas in Tellus.

the problem or problems of concern. The next step is to use whatever general information on the environment is available (yes, our past efforts at environmental measurement have some use) to develop alternative hypotheses giving possible solutions to the problem as stated. Each hypothesis then provides a set of questions that we must ask of the environment in order to prove or disprove the subject hypothesis. The data-collecting program should then be designed to answer each of the individual questions required to prove or disprove the formulated hypothesis.

Admittedly, there may be some areas of environmental study for which so little general knowledge is available that no reasonable hypothesis may be formulated. I feel that this situation would be exceptional, for if we know enough to clearly state the problems that need solution, we must know something of the environment, if only by analogy to similar, better-studied situations.

In essence, our starting point should be a set of conclusions, and our purpose should be to find which of these conclusions are most nearly correct and which are clearly not correct. We must know what techniques of data interpretation are available, and which would most clearly serve our specific purpose. We then can design an observational program that will supply data compatible with the interpretation techniques we have selected.

In my presentation here, I have departed from the proposed content of my paper as stated by the organizers of this meeting in the printed symposium program. I have not given any gems of knowledge about trends and cycles, cause and effect relationships, statistical inferences, or direct and indirect conclusions that will greatly help any of you interpret the mass of generally inadequate existing environmental data. What I have tried to do is to present some concepts that I hope might be employed in the development of the new extensive and expensive environmental studies now being planned or contemplated.

The modern statistical methods of treating environmental data, such as power spectrum analysis, have been adequately discussed in the literature, and I only hope that other speakers at this symposium will have given some general information on what interpretations and conclusions can be drawn from their use. In the time allotted I must be satisfied (even if my listeners are not) with this broad statement of the philosophy that should be pursued in future environmental studies.

Dr. Leslie A. Chambers
Professor of Biology and
Director, Allan Hancock Foundation
University of Southern California, Los Angeles

SUMMATION

Some of us can recall occasions, only a few decades back, of conferences on technical and scientific subjects wherein the objective was brief communication of new ideas and findings, informal discussion of their signicance, and a comforting absence of preprints. Very gradually, and in parallel with the ready assumption by scientists and engineers of a new order of economic and social respectability, group communication among us has become stylized to an astonishing degree. Now there are indices of status built into every conference — indices which stratify the convening agency and the conferees themselves much more certainly than the informational content of the session itself.

Instead of contributed papers we now have symposia of invited speakers on prescribed subjects; instead of concise introduction of contributors by name and title of his paper, we now have lengthy accolades listing past honors, achievements, and other biographical notes; instead of a prompt entry into the subject area of the symposium, we now invariably have an hour or two or more of successive introductions and welcomes culminating in the expected words from the highest-ranking individual the conveners have been able to woo away from his normal duties. To cap the procedure, to endow it with the formal attributes of stature, some near-pensioner, formerly but not now active in the general area of conference purview, is customarily enlisted to say a blessing over the whole thing in the guise of a "summation." I am honored by this role this morning, but have never before sensed so fully the non-essentiality of a symposium summary.

Those of you who have sat with me through the general sessions and a selection of the separate subsessions dedicated to water and air will easily recognize the dilemma. How can one possibly abstract a set of abstracts, epitomize an encyclopedia, minimize a minimum. The enormous breadth of our subject area — environmental measurements — coupled with the extraordinarily successful efforts of the several speakers to compress their assigned facets of the whole into a few minutes, has given birth to what, when published, will be a kind of pocket reference manual in the philosophy, technology, and symbolism of communications theory, experimental design, statistical operations, machine analysis computer programming, and a variety of other more or less related concepts. It would be an injustice to some of the excellent papers to squeeze them further or to take items from them; certainly no purpose can be served by offering orally an annotated index of speakers and titles. You have the flavor of the conference sessions, your personal estimates of the several contributions, and you have the papers themselves to read and re-read if you are intrigued.

All of this leads to the simple fact that I have no intention of attempting any summarization paper by paper or session by session. With apologies to the individuals who have contributed, but without specific acknowledgment of their respective contributions, I shall instead, use the next five minutes to summarize my own reaction to the conference as a whole, and to add a comment or two in the philosophical vein so ably mined by Dr. Anderson on Wednesday.

The papers we have heard fall into three general categories: (1) those that dealt with the philosophy of measurement, information transfer, and data interpretation, (2) those that offered in didactic but often delightful fashion certain elementary principles of statistical theory, experimental design, and computer programming, and finally, (3) a considerable number that reflected their author's preoccupation with the application of measurement techniques and analyses to specific problem objectives. The third category has tended to exemplify the inadequacies of the existing theories, techniques, sensing equipment, and concepts, *or more probably* has exemplified a crying need for more skill and understanding in their application to concrete problems of environmental measurement.

Some of the questions that arise in any study of the environment have been asked here, and some have been answered in part.

Why do we measure? The importance of a clear understanding of the objective has been emphasized; in the lingo of experimental science, a clear and understood statement of the problem to be solved is the prime requisite. Even at this point, a group of people such as this will certainly formulate different starting points. Those with an end-point perspective — the engineers, and physicians for example — will measure an environmental parameter for the purpose of future interpretation in terms of some possible effect on man or other object. Those who find their pleasure in sheer understanding of the properties of a given system may be content with information about the system for its own sake. In either event, they all qiuckly find that the classes, substances, or events they have chosen to measure, even if completely mensurable, cannot by themselves give any final satisfaction. The man whose objective is to hold the environment in compatibility with human tolerance must include himself and other men as reactants in the system he considers. If, for example, the concern is with lead in the atmosphere there can be no useful result from measurement of airborne lead alone; there must also be data on the ranges of human tolerance to lead as functions of age, rate of intake, physical and chemical forms in which the lead occurs, and especially there must be data on lead intake in water and food and the relative importance of intake by different routes.

If, on the other hand, the concern is with airborne lead as part of the normally dynamic atmospheric system per se, its role cannot be interpreted from static measurements of lead concentrations alone. Interactions are the norm and their products may not even be precisely definable as lead in a proper sense.

Such considerations lead promptly to a series of additional questions.

What should we measure in order to attain the defined objective? Certainly airborne elemental lead values will prove insufficient for almost any purpose.

How should we measure the several parameters essential to attainment of our special objective?

How much measurement is necessary? In other words, what is the minimum effort necessary to achieve some reasonable level of significance taking into account the accumulative errors of the several types of measurement involved?

When should the measurements be made? Is the effect of the man-environment interaction expected to be a long or short function of time?

How do we process the data to produce a display of interpretable evidence bearing on the pre-set problem?

And finally, *how* do we communicate the evidence and findings to create maximal momentum toward a control objective?

We came here with these questions before us; we leave with the certainty that there is no pre-mixed formula that will permit transfer of our central functions to the best of present or future sensing, data processing, and analytical labor-saving systems. It is important to bear in mind that the thinking necessary to the programming of the finest systems may be the weakest link in the chains of events we set in motion. On the other hand we leave impressed by the rapidity with which cybernetic extensions of our inherent capabilities are enabling some reasonable approaches to environmental problems involving multiple parameters.

At the conclusion of his paper, Gaylord Anderson drew from the Homeric version of the Straits of Messina a classic allegory in which the cooperation of Scylla and Charybdis absorbed an input of fragmentary evidence and imprecise data and spewed forth false conclusions. In an earlier portion of the same paper attention was called to the fact that even with attention focussed strongly on physical and chemical factors in the environment, one must not forget that man and his behavior have produced the environmental alterations which we fear. And elsewhere pointed out that the basic reason for measurement is to determine the magnitude of environmental forces and the effects they have on man.

Now, the minor logical conflict I read in these two statements (taken out of context) I find it possible to resolve quite readily. There is no *a priori* necessity for regarding either man or his environment as either cause or effect. In a very real sense man is simply another reactant in the nonhomogenous, many-parametered system with which we are concerned. It is a single, temporally continuing, constantly interacting system with which we deal. It is with variations in rates and quantities that we are concerned since the system, with or without rather transient intermediate steady-state assemblages such as man, has been here a very long time, and will be here much longer. The time has almost come, in terms of operational capability, when we can begin to think of the simultaneous analysis of entire segments of the system and not be restricted to adaptive actions based on measurements of single parameters plus intuition, hope, and a graveyard rabbit's foot.

If we are to tackle analysis of the system of which we are a part, a level of technical skill, mathematical and logical sophistication, and philosophical detachment not now generally incorporated into the training of environmental scientists will have to be attained. This symposium has made its contribution; as realization of the total requirements to cope with the total problem is attained, succeeding conferences will undoubtedly be more comprehensive in scope and at the same time, deal in greater depth with the technologoy of planning, sensing, transmitting, translating, analyzing, and displaying essential information.

7

SESSION 7: Measurements of Air Environment

Chairman: **Jean J. Schueneman**
Chief, Technical Assistance Branch
Division of Air Pollution
U. S. Public Health Service

Dr. Ralph I. Larsen
Field Studies Branch
Division of Air Pollution
U. S. Public Health Service, Cincinnati

SUMMARY

Interrelations among variables may be determined by the following steps. Plot the data. Study the variables that show good interrelationships. Determine if the interrelation is arithmetic, semi-logarithmic, logarithmic, cyclic, or probabilistic. Plot the data on a type of graph paper that will give a straight line. Determine the equation of the line, thus tersely expressing the relationship between variables. Correlate and regress the data. Construct and test a mathematical model that agrees with the results and makes good physical sense. Try to understand and explain why the relationship exists. Use the new knowledge gained to better manage the environment, whether it be air, water, land, radiation, milk, food, or something else.

DETERMINING BASIC RELATIONSHIPS BETWEEN VARIABLES

"For all the Athenians, and strangers which were there, spent their time in nothing else but either to tell or to hear some new thing."[1]

Thus began the response of Dr. Joel H. Hildebrand upon receiving the 1962 William Procter Prize from the Society of the Sigma Xi.[2] He continued with, "All true scientists like to spend most of their time in nothing else than either to discover, to tell, or to hear, some new thing.

"The urge begins with a peculiar combination of genes which produces an insatiable curiosity. This leads in childhood to endless questions and continual experiments with things and persons. The behavior is not that of the model 'good child,' who, when told to run along and not to ask so many questions, obediently 'runs along,' never asks questions his elders cannot answer."

He continues by noting, "A physical scientist does not merely 'learn' the laws of thermodynamics; he must try to understand them; he must gain an intuitive feeling for the concepts of enthalpy, energy, free energy and entropy . . . Even so delightful a subject as calculus can be taught mainly as formulas for differentiating and integrating, whereas what is really needed is that a person shall understand the various expressions and operations so well that one can formulate a physical problem in mathematical terms, translating freely back and forth between English and calculus."

This curiosity and desire for understanding noted by Dr. Hildebrand are two important motivating forces needed for exploring basic relationships between variables.

THE PROBLEM

Today's environmental studies produce thousands and sometimes millions of numbers. The desire to understand their meaning forces one to try to determine their interrelationships.[3] Ideally, the results should distill into a few cogent formulae, just as Newton distilled his observations into three laws of motion, his second one being

$$\text{force} = (\text{mass})\ (\text{acceleration})$$

Similarly, Einstein conducted almost no experiments of his own, but used the results of others to formulate his theory of relativity, and his world-shaking

$$\text{energy} = (\text{mass})\ (\text{speed of light})^2$$

APPROACH

The thoughts that follow are from my own limited experience. Others might emphasize different points.

Insight into possible new relations between variables or new analytic approaches to a study seldom comes when I am working hard directly on the study. Instead it comes when I may be tugging some thoughts over in my mind rather loosely and relating things from different fields. It is in this atmosphere that new approaches may come to mind. Also, for myself it seems to work best to do the hardest or most demanding or creative work in isolation before noon, and maybe communicate and do more routine work after noon.

So much for attitudes and philosophy. What methods can be used to explore and determine basic interrelations between variables?

METHODS

The following sequence of operations works best for me:

1. Plot tens or hundreds of plots of one variable or group of variables against the others. Use simple cartesian coordinate paper. Plot with pencil, punched card tabulator,[4] or electronic computer.[5]

2. Study in detail the plots that indicate good relationships, i.e., without widely scattered points.

3. Determine if the data are cyclic. If not, find a graph paper on which the data will plot as a straight line.

4. Determine the equation of the line, thus tersely expressing the relationship between variables.

5. If you want to find out how good the relationship is, correlate and regress the values. If you have quite a few values, let an electronic coputer do this.[4, 5]

6. Construct a mathematical model that agrees with the results and, preferably, makes good physical sense as well. Test the model.

7. Try to understand and explain why the relationship exists.

8. Use the new knowlege gained to better manage the environment, whether it be air, water, land, radiation, milk, food, or whatever.

EXAMPLES

Let us consider examples of various sets of data and various types of interrelations.

ARITHMETIC

Baulch has recently related sulfur dioxide concentration to wind direction gustiness.[6] He classified gustiness into five types and showed distributions of the time percent of each type for various sulfur dioxide concentrations. Gustiness types B_1 and D seemed to be especially related to concentration. They are defined as follows:

B_1 : Wind fluctuations from 15 to 45 degrees.

D : Short-term fluctuations not exceeding 15 degrees. The trace approximates a straight line.

The ratio D/B_1 looks as if it should relate to sulfur dioxide concentration. Three of the points plot as a straight line on cartesian coordinate paper (Figure 1). The equation of a straight line is

$$y = mx + b$$

Where m is the slope of the line and b is the value of y at x = 0.

Thus the equation for Figure 1 is

$$\text{sulfur dioxide conc.} = 0.033 + 0.02 \left[\frac{D}{B_1} \right]$$

(Actually, if all four points were considered, a semi-logarithmic plot would fit best. An arithmetic plot is used here for an example.)

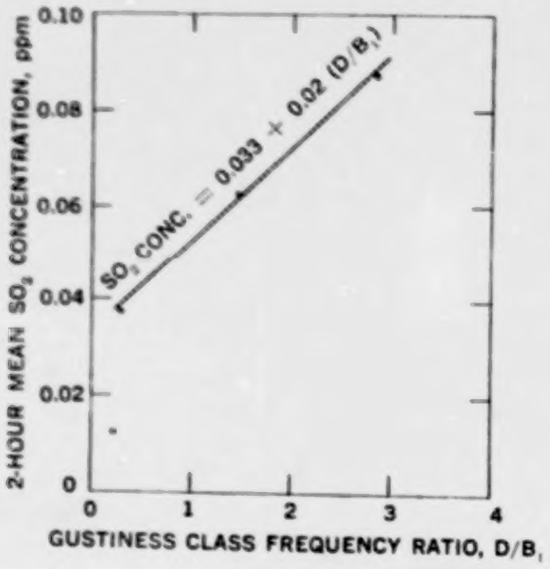


Figure 1 — Two-Hour Mean Sulfur Dioxide Concentration in Nashville Versus Wind Gustiness Ratio, October 1958 - March 1959. (Data Source: Ref. 6).

Gustiness type D indicates stable meteorologic conditions; type B_1 indicates more turbulent conditions. Thus the equation makes sense in that stability is associated with high concentrations and turbulence is associated with low concentrations.

SEMI-LOGARITHMIC

Tice has presented steel corrosion data as a function of time exposed (Figure 2).[7] These data look is if they would fit a straight line if the lower years were expanded to the left. A logarithmic horizontal scale will accomplish this (Figure 3). The equation for this line may be determined as follows. Again, the equation of a straight line is

$y = mx + b$

x will be on a logarithmic scale

$y = m(\log x) + y \text{ at } \log x = o$

$$m = \frac{y_2 - y_1}{\log x_2 - \log x_1}$$

$$= \frac{84 - 36}{\log 10 - \log 1} = \frac{48}{1{\cdot}0} = 48$$

$y = m(\log x) + y_{x = 1}$

$y = 48 \log x + 36$

Weight loss, in g = 48 log (years) + 36

Similar methods can be used to determine equations for the other two lines. Reducing the data to equations allows describing any line by only two parameters, slope and intercept. These two parameters may then be used to compare plots.

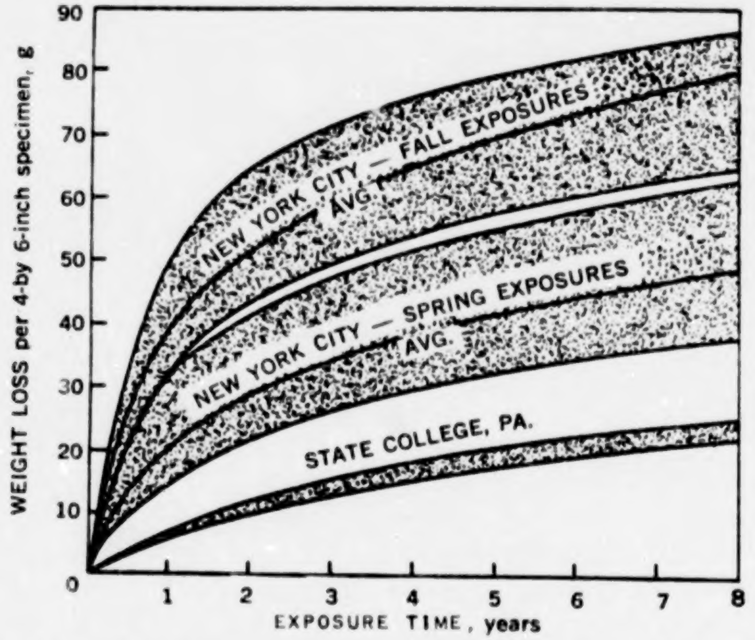


Figure 2 — Effect of Time of Start of Tests on Corrosion of Steel at New York City. (Source: Ref. 7).

This semi-logarithmic plot indicates that the same weight loss occurs with every doubling of time. Thus the same weight loss occurs between 1 and 2 years as between 2 and 4 and 4 and 8. As corrosion occurs, fewer open sites are left to corrode, and oxidation below the oxide layer is probably slower. This might be one possible explanation for a decreased corrosion rate as a function of time.

This is an example of a logarithmic horizontal plot. Let us now consider a logarithmic vertical plot, and one with three variables instead of two. You can disregard the parameters and think only of the mathematics. The parameters happen to be salary versus Government Service grade and step (Figure 4). It appears that a straight line might result if the lower grades could be expanded downward. A logarithmic vertical scale would accomplish this (Figure 5). A definite change in slope occurs at GS 11. Thus one line could describe GS 1 to 11 and another GS 11 to 15.

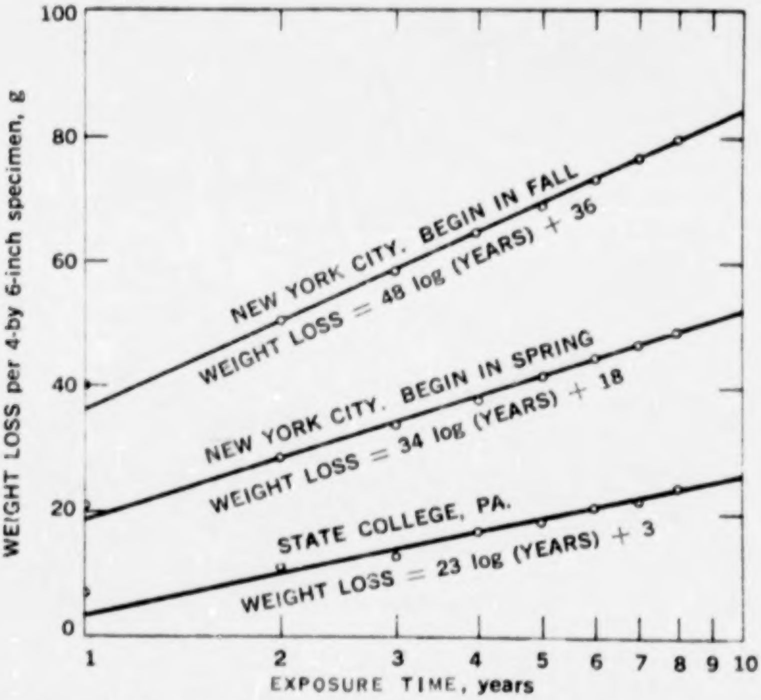


Figure 3 — Weight Loss of Steel Versus Years of Exposure. (Data Source: Ref. 7).

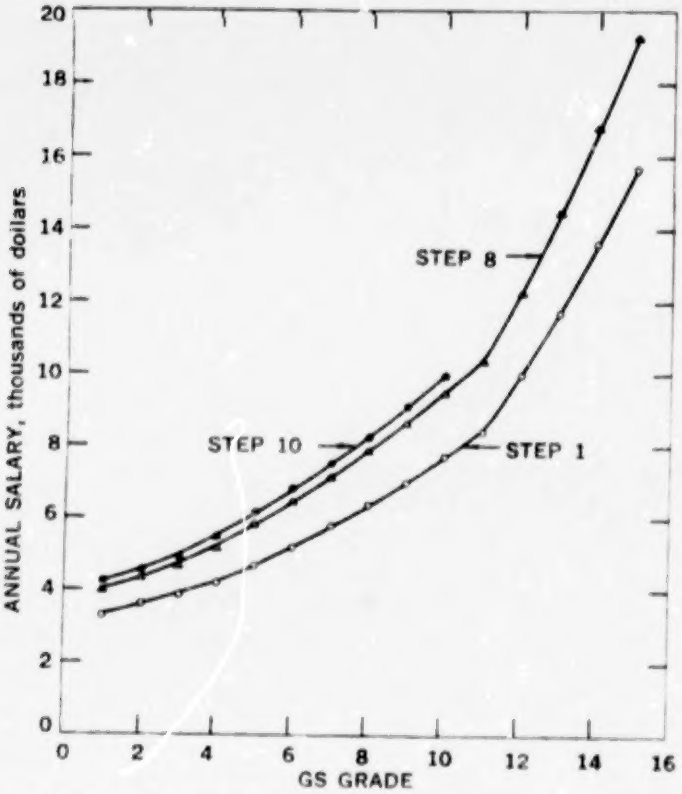


Figure 4 — Annual Salary Versus Government Service Grade, January 1, 1964.

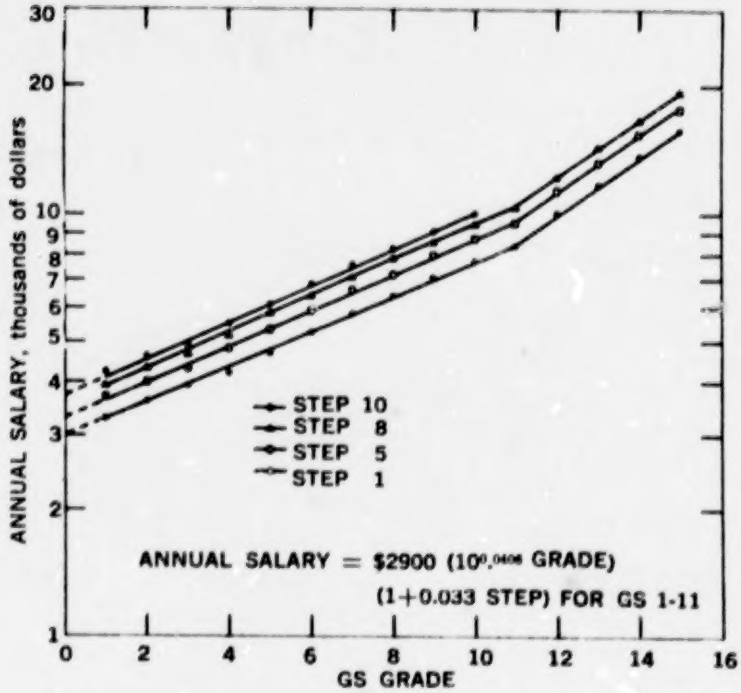


Figure 5 — Annual Salary Versus Government Service Grade, Logarithmic Vertical Scale.

Again, for a straight line

$$y = mx + b$$

In this case, y is on a logarithmic scale

$$\log y = mx + \log y \text{ at } x = 0$$

$$m = \frac{\log y_2 - \log y_1}{x_2 - x_1}$$

$$= \frac{\log 7{,}650 - \log 3{,}000}{10 - 0}$$

$$= \frac{3.884 - 3.478}{10} = \frac{0.406}{10} = \underline{0.0406}$$

$$\log y = 0.0406\ (\text{GS grade}) + \log \$3{,}000$$

Take antilogs

$$\text{Annual salary} = \$3000\ (10^{0.0406\ (\text{GS grade})}) \text{ for GS } 1 - 11$$

$$\text{Annual salary} = \$1500\ (10^{0.0682\ (\text{GS grade})}) \text{ for GS } 11 - 15$$

Step 10 rates plot parallel to step 1 rates (Figure 5). Thus the vertical distance from step 1 to step 10 is a constant. For this logarithmic scale, step 10 is thus always 30 percent greater than step 1, regardless of the grade.

A plot of salary as a function of step is linear (Figure 6), indicating that for a given grade each step is a constant number of dollars greater than the previous step.

Combining the effects of grade and step, salary may be expressed as follows:

$$\text{Annual salary} = \$2900\ (10^{0.0406\ \text{grade}})\ (1 + 0.033\ \text{step})\ \text{for GS 1-11}$$

$$\text{Annual salary} = \$1450\ (10^{0.0682\ \text{grade}})\ (1 + 0.033\ \text{step})\ \text{for GS 11-15}$$

Since three variables are involved, the data may be plotted in three dimensions using isometric paper. The equations describe the top surface of Figure 7.

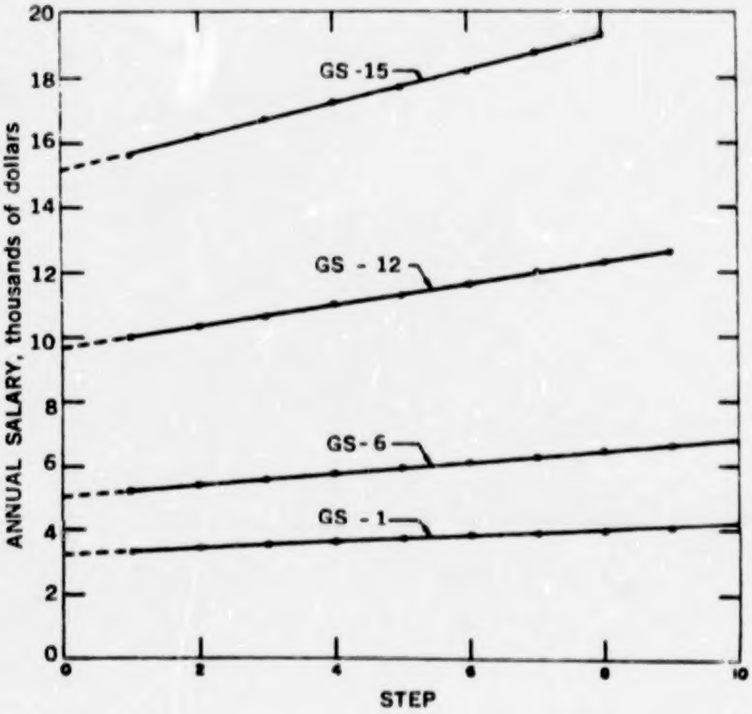


Figure 6 — Annual Salary Versus Step for Various GS Grades, January 1, 1964.

Possibly an easier way to think of the data is that each grade, from 1 to 11, pays 10 percent more than the previous one; and each grade from 11 to 15 pays 17 percent more than the previous one.

The salaries were probably not determined in this manner, since they do vary above and below the trend, but the equations seem to give a good estimate of the interrelations between grade, step, and salary.

LOGARITHMIC

On cartesian graph paper, if one end of a plot tends to become horizontal and the other end vertical, a logarithmic relation may exist. A plot of the percent oxy-hemoglobin in the blood as a function of the partial pressure of oxygen seems to satisfy this requirement (Figure 8).[8] To fit a logarithmic plot, however, the graph has to be turned upside down and percent arterial hemoglobin unsaturation used rather than saturation. A straight line then results (Figure 9).

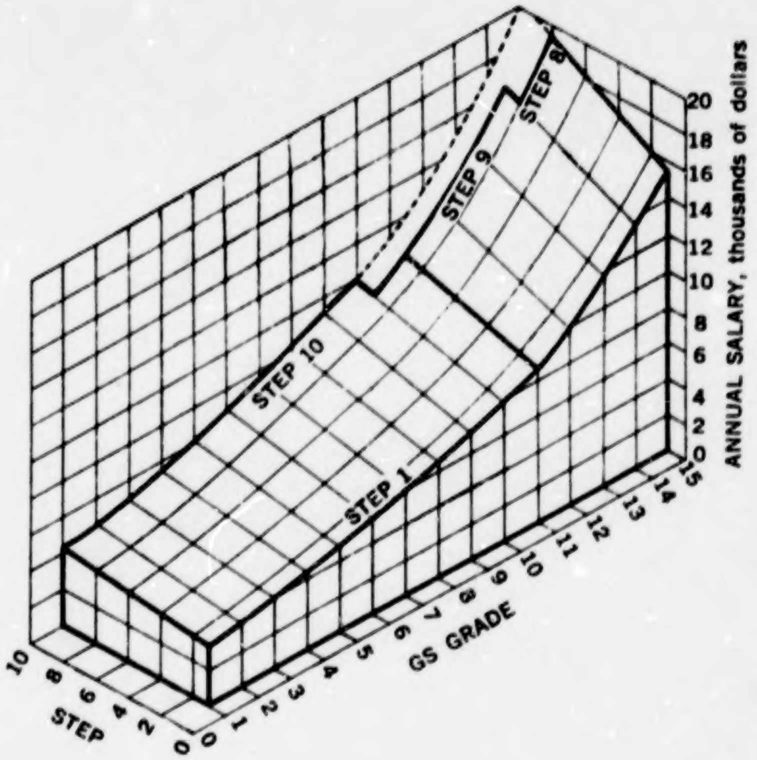


Figure 7 —Annual Salary as a Function of Government Service Grade and Step, January 1, 1964.

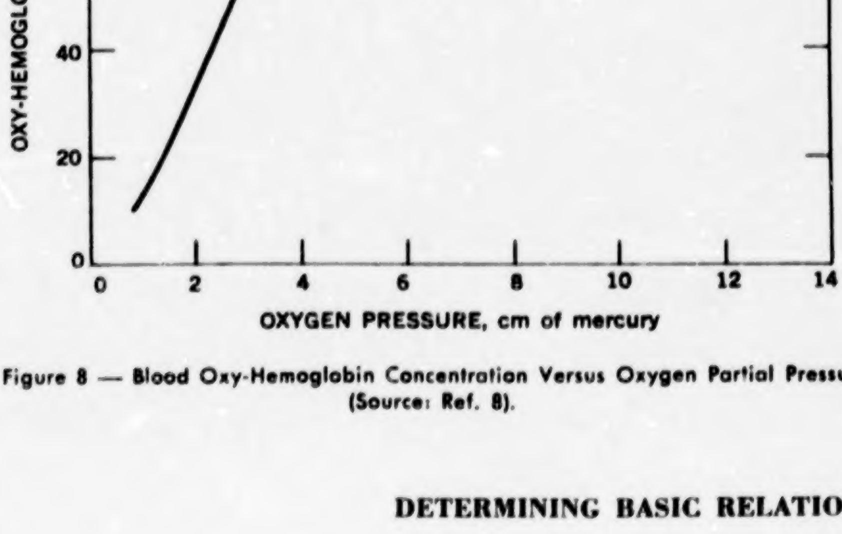


Figure 8 — Blood Oxy-Hemoglobin Concentration Versus Oxygen Partial Pressure. (Source: Ref. 8).

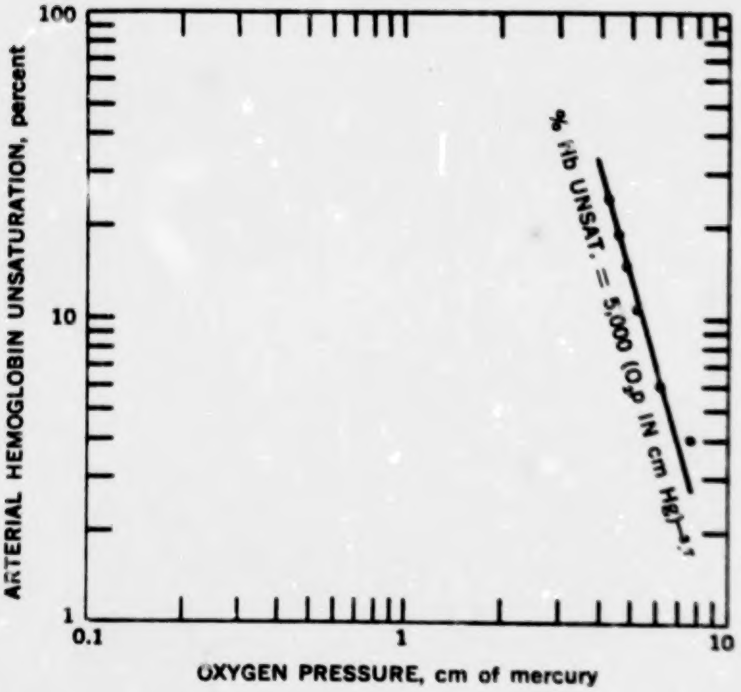


Figure 9 — Percent Oxy-Hemoglobin Unsaturation Versus Oxygen Partial Pressure. (Data Source: Ref. 8).

Again, for a straight line

$$y = mx + b$$

In this case both axes are logarithmic.

$$\log y = m(\log x) + \log y \text{ at } \log x = 0$$

$$m = \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1}$$

m may also be determined by merely measuring the slope on the graph with a scale or ruler.

Take antilogs

$$y = b\,x^m$$

b is the value of y at log x = 0 (i.e., x = 1).

Thus for Figure 9, since y is 5,000 when x = 1,

$$\%\text{ hemoglobin unsat.} = 5{,}000\ (\text{oxygen pres. in cm of mercury})^{-3.7}$$

This indicates that hemoglobin unsaturation is inversely proportional to oxygen pressure to almost the 4th power.

PROBABILITY

The distribution plot of many entities in the world is bell-shaped, fitting an arithmetic-probability or Gaussian distribution. Air pollutant concentration data usually fit this bell shape, if concentration is plotted on a logarithmic scale,[9,10] giving a logarithmic-probability plot (Figure 10).[4] To get a straight line, distributions of data may be plotted on cumulative distribution paper, either arithmetic-probability or logarithmic-probability, whichever fits a straight line best.

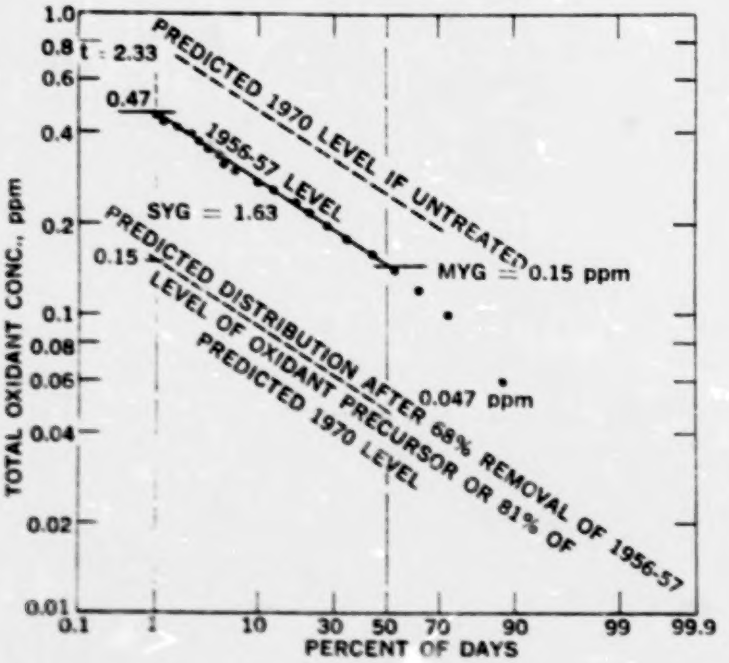


Figure 10 — Frequency of Various Levels of Total Oxidant Peak Hourly Concentration at Los Angeles Station 1, 1956-57.

A plot of sulfation (an index of sulfur dioxide concentration) as a function of the distance from the center of Nashville showed this typical bell-shaped distribution (Figure 11).[11] Plots of the data on arithmetic cumulative distribution paper gave straight lines (Figure 12). Therefore it was possible to use a Gaussian-type equation to express sulfation in Nashville as a function of the distance from the center of town.

$$S = S_b + S_c e^{-\frac{1}{2}} \left(\frac{r}{s_r}\right)^2$$

where S is sulfation,

S_b is the background sulfation,

S_c is sulfation at the center of Nashville (minus S_b),

e = 2.718, the base of natural logarithms,

r is the radial distance from the center of Nashville, and

s_r is the standard radial deviation, which is analogous to standard deviation.

CYCLES

Some variables vary cyclically. Oxidant concentration in Los Angeles is a function of sunlight, and thus tends to peak about noon and be low at night. Concentrations of pollutants from motor vehicles tend to peak during the morning and evening traffic rushes and be lower at other times. Many variables are a function of time of day, day of week, season, year, or maybe even sun spot intensity (11-year cycle). These variables may be expressed as sine or cosine waves with none or several harmonics. In fact any continuous curve may be approximated by a sufficient number of harmonics, by means of Fourier analysis.[12] More intricate time series techniques may also be used for auto correlation and power spectrum analyses.

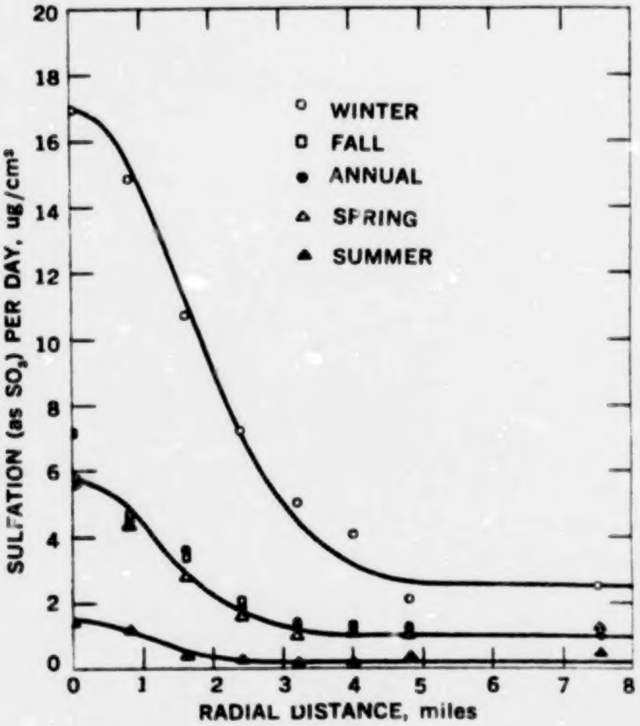


Figure 11 — Geometric Mean Sulfation by Season Versus Radial Distance from Center of Nashville.

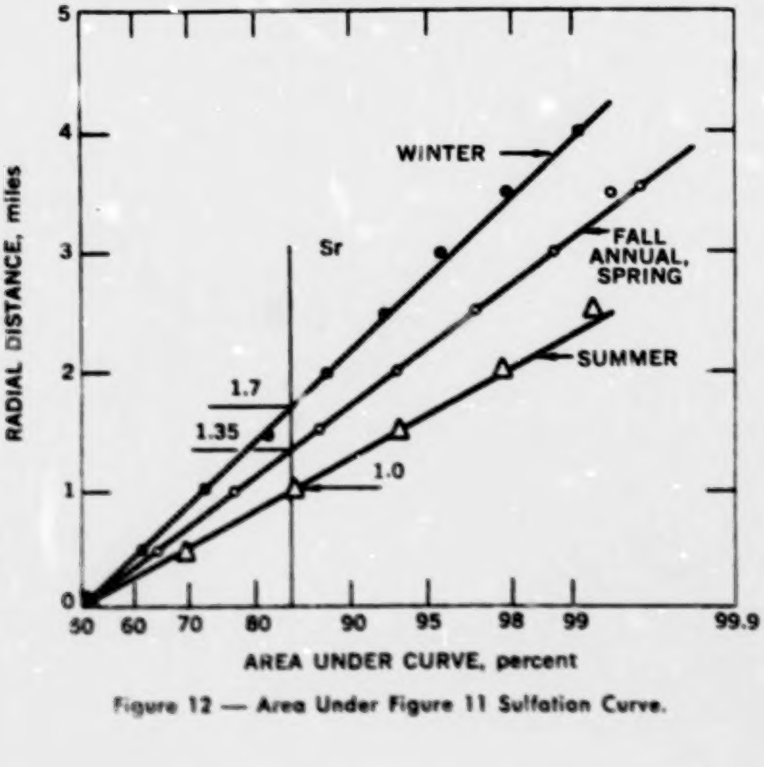


Figure 12 — Area Under Figure 11 Sulfation Curve.

A time plot of minimum sun spot activity and air pollution disasters is interesting (Figure 13). Four air pollution disasters have occurred during the past three peace-time periods of minimum sun spot activity, including the period we are in presently. The Donora disaster is the exception. Whether a real interrelation exists or whether this is merely happenstance, I do not know, but it is interesting to contemplate, and possibly to predict, "Look out for the winters of 1962-64."

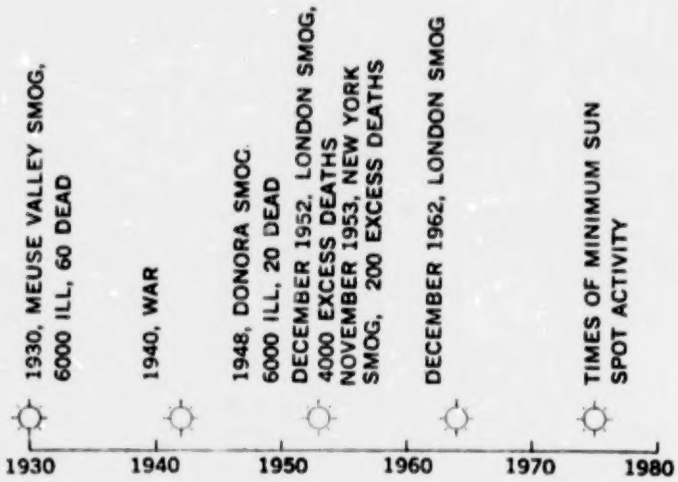


Figure 13 — Air Pollution Disasters and Minimum Sun Spot Activity, 1930-1980.

MATHEMATICAL MODELS

Mathematical models for explaining data can be proposed and then checked for validity.[13] We have just seen in the Nashville example how sulfation can be expressed. Sulfur dioxide emission data can be expressed in a similar manner. Sulfation can then be related to emission. A simple mathematical model was proposed to do this.[11]

$$\text{sulfation} = \frac{k(\text{emission strength})}{x^n}$$

where k is a constant,

x is the distance between source and receptor, and

n is an exponent.

The problem was programmed on a computer and tested for several combinations of parameters. The best fit to actual data occurred with n = 2, indicating that sulfation is inversely proportional to the square of distance between source and receptor. This makes sense, for it indicates that the long-term average diffusion of pollutants from a point is similar to the diffusion of light or the diffusion of nuclear radiation from a point. Or maybe it would be better to say that the radiation of air pollutants is similar to the radiation of other mass or energy.

The climax of many studies is building, testing, and validating mathematical models to explain the data. Ideally, the model should be a mathematical expression of the actual physical process involved, or a simplified representation of the process. The model might describe the interactions between mass and energy in air, water, man, and bacteria. Algebra, calculus, and statistics could all be used in constructing the model.

Model construction and validation may be the most vital and challenging part of a study. Unfortunately, interest in a study may flag by this time, or preparation for the next study may be demanding. Thus this vital, key operation may be seriously neglected. It is important, however, to spend lots of time in thought, cogitation, and testing at this stage, in order to produce a finished product.

REFERENCES

1. Luke, "Acts of the Apostles," *Bible*, Acts 17:21.

2. J. H. Hildebrand, "To Tell or to Hear Some New Thing," *American Scientist*, 51:2-11 (March 1963).

3. R. I. Larsen, "Parameters of Aerometric Measurements for Air Pollution Research," *American Industrial Hygiene Association Journal*, 22:97-101 (April 1961).

4. R. I. Larsen, "A Method for Determining Source Reduction Required to Meet Air Quality Standards," *J. Air Poll. Control Assoc.*, 11:71-76 (February 1961).

5. R. I. Larsen, "Choosing an Aerometric Data System," *J. Air Poll. Control Assoc.*, 12:423-430 (September 1962).

6. D. M. Baulch, "Relation of Gustiness to Sulfur Dioxide Concentration," *J. Air Poll. Control Assoc.*, 12:539-542 (November 1962).

7. E. A. Tice, "Effects of Air Pollution on the Atmospheric Corrosion Behavior of Some Metals and Alloys," *J. Air Poll. Control Assoc.*, 12:553-559 (December 1962).

8. R. A. McFarland, F. J. W. Roughton, M. H. Halperin, and J. I. Niven, "The Effects of Carbon Monoxide and Altitude on Visual Thresholds," *J. of Aviat. Med.*, 15:381-394 (1944).

9. C. E. Zimmer, E. C. Tabor, and A. C. Stern, "Particulate Pollutants in the Air of the United States," *J. Air Poll. Control Assoc.*, 9:136 (November 1959).

10. *Air Pollution Measurements of the National Air Sampling Network, 1957-1961*, Public Health Service Publication 978, U. S. Government Printing Office, Washington, D. C. (1962).

11. R. I. Larsen, W. W. Stalker, and C. R. Claydon, "The Radial Distribution of Sulfur Dioxide Source Strength and Concentration in Nashville," *J. Air Poll. Control Assoc.*, 11:529-534 (November 1961).

12. H. A. Panofsky and G. W. Brier, *Some Applications of Statistics to Meteorology*, The Pennsylvania State University Press, University Park, Pennsylvania (1958).

13. E. K. Harris, D. S. Licking, and J. B. Crounse, "Mathematical Models of Radionuclides in Milk," *Public Health Reports*, 76:681-690, (August 1961).

Glenn W. Brier
Chief, Meteorological Statistics Research Project
U. S. Weather Bureau, Washington, D.C.

SUMMARY

Because values in a time series may not be statistically independent, the reliability of various statistics generated from time-series data may be questioned. The tendency of each value to be correlated with chronologically adjacent values is known as persistence, a problem that requires the application of special methods. A procedure for spectral estimates and a filtering or smoothing function are applied to the analysis of meteorological data. The significance of high-speed computer technology is emphasized.

INTERPRETATION OF TRENDS AND CYCLES

During the past few days you have heard of a number of statistical concepts or principles and have been introduced to a few techniques of data analysis. A set of n observations

$$x_1, x_2, x_3 \cdots\cdots x_n$$

has been treated or analyzed as a sample from a "population" by some appropriate theory that makes use of a mathematical or probabilistic model. The usual assumption is that the n observations are independent — that one actually has a sample of n observations. In much meteorological or geophysical data, however, the value of a particular x_1 is not statistically independent of the other values in the sample and may be related to x_1+_1 or x_1+_2, for example, because of proximity in space or time. This interdependence of values tends to invalidate the standard formulas used to assess the reliability of the various statistics estimated from the data, such as means, standard deviations, correlation coefficients, etc. In a time series, as a rule, the successive values of the series are not independent of one another, and the tendency of each value to be correlated with chronologically adjacent values is known as persistence. Special methods are needed to treat this problem of persistence in data; today we will consider a few of the things that might be done. Some aspects of the problem have been reviewed and discussed recently by Mitchell.[6]

One of the oldest questions in meteorology is whether there are any cycles in weather data, other than the well-known daily and annual cycles. This question has been investigated by hundreds, who have used the classical methods of harmonic analysis known to mathematicians for centuries. This technique is a proper one for investigating the harmonics of a fixed identifiable frequency under the assumption that the time series is genuinely periodic, i.e., repeats itself exactly every n observation. Its misuse "when these assumptions do not hold has been responsible for the acceptance of probably more spurious hypotheses than any other statistical or applied mathematical tool . . . [It] breaks down completly when applied to a statistical fluctuation."[5]

Now if harmonic analysis is not an appropriate tool for use in investigating nonrandomness and apparent quasi-periodic fluctuations in data, what can we use? One answer is that given by Tukey,[8] who suggested a sound and practical computational procedure for obtaining "spectral" estimates based on the results of the pioneer work by Wiener[10, 11] on generalized harmonic analysis.

The recommended procedure provides spectral estimates (U_h) showing how the variance of the time series is distributed as a function of frequency. Ward[9] has recently described the method in some detail in connection with an application to the geomagnetic disturbance indices. Panofsky and Brier[7] discuss some other applications, and Blackman and Tukey treat the subject more completely in a recent monograph.[1]

The spectral estimates are obtained by first computing the sample autocorrelation function

$$R_k = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i \, x_{i+k}$$

where n is the number of observations used and the x_i are data points expressed in terms of the deviation from the mean of the series. From the Fourier-transform of the R_k function, the apparent "line powers" L_h are determined by

$$L_o = \frac{1}{2M}(R_o + R_M) + \frac{1}{M} \sum_{K=1}^{M-1} R_k$$

$$L_h = \frac{1}{M}(R_o + [-1]^h R_M) + \frac{2}{M} \sum_{K=1}^{M-1} R_K \cos \pi K \frac{h}{M}$$

$$L_m = \frac{1}{2M}(R_o + [-1]^M R_M) + \frac{1}{M} \sum_{K=1}^{M-1} (-1)^K R_K$$

M is the number of lags for which the autocorrelation function is computed and is usually about 5 or 10 percent of the number of observations n.

The values of L_h are smoothed to obtain the spectral estimates

$$U_o = 0.54\,L_o + 0.46\,L_1$$

$$U_h = 0.54\,L_h + 0.23(L_{h-1} + L_{h+1})$$

$$U_M = 0.54\,L_M + 0.46\,L_{M-1}$$

Tests have been given that enable one to determine whether a spectral peak departs significantly from some specified base line, such as that expected from a flat or "white" noise spectrum. The white noise spectrum is one in which all frequencies contribute equally to the total variance of the series, and would be expected if a set of random numbers were analyzed, for example. Figure 1 shows the spectral estimates obtained from the analysis of 110 years of annual precipitation values for Copenhagen, Denmark. Although these data were analyzed in connection with an interest in a possible 11-year sunspot cycle, there is no statistically significant peak near 11 years nor anywhere in this spectrum.

Although power spectrum analysis has been found to be a valuable tool in the study of time series, it is limited in application and often should be supplemented by other types of analysis. Spectrum analysis discards phase information as well as the details of any amplitude variation. Sometimes it may be desirable to recover this information to gain a little insight into what is going on in the original series. One way of accomplishing this is by smoothing or filtering. In these techniques the original series x_t is operated on by a "filtering function," or perhaps by several such functions. These methods have been discussed by Holloway,[4] Panofsky and Brier,[3] and others. The simplest commonly used method is to eliminate or reduce the amplitude of the short-period fluctuations or "noise" by the use of running averages. This is a special case of the general procedure of treating the observations x_t in the time series by the following linear equation

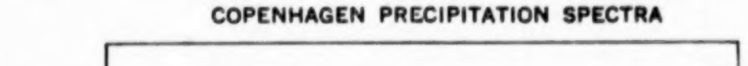

COPENHAGEN PRECIPITATION SPECTRA

Figure 1 — Spectral Estimated Relative Power (U_h) of the Annual Precipitation Totals for Copenhagen, Denmark.

$$F_t = \sum_{K=-n}^{M} W_K x_{t+K}$$

where W_K is a particular weight in the filtering function. The weight W_0 is known as the principal weight or the central weight when the filter is symmetrical with $n = M$. In the process of filtering the time series, successive observations are cumulatively multiplied by these weights, producing a new series beginning F_t, F_{t+1}, F_{t+2} and continuing in this succession.

A filter that reduces the amplitude of both the high- and low-frequency fluctuations, leaving a middle range of frequencies relatively unaffected, is called a band-pass filter. Such a filter is useful in studying the fluctuations of a particular time scale. For example, Table 1 gives the weights used for a filter having the maximum sensitivity to fluctuations

Table 1 — Set Of Weights Used For Band-pass Filter

W-27	0.0100	W-0	0.0602
W-26	0.0121	W-1	0.0577
W-25	0.0140	W-2	0.0512
W-24	0.0156	W-3	0.0410
W-23	0.0165	W-4	0.0279
W-22	0.0163	W-5	0.0133
W-21	0.0147	W-6	—0.0017
W-20	0.0115	W-7	—0.0158
W-19	0.0065	W-8	—0.0279
W-18	0.0000	W-9	—0.0372
W-17	—0.0079	W-10	—0.0431
W-16	—0.0165	W-11	—0.0454
W-15	—0.0253	W-12	—0.0442
W-14	—0.0334	W-13	—0.0400
W-13	—0.0400	W-14	—0.0334
W-12	—0.0442	W-15	—0.0253
W-11	—0.0454	W-16	—0.0165
W-10	—0.0431	W-17	—0.0079
W-9	—0.0372	W-18	—0.0000
W-8	—0.0279	W-19	0.0065
W-7	—0.0158	W-20	0.0115
W-6	—0.0017	W-21	0.0147
W-5	0.0133	W-22	0.0163
W-4	0.0279	W-23	0.0165
W-3	0.0410	W-24	0.0156
W-2	0.0512	W-25	0.0140
W-1	0.0577	W-26	0.0121
		W-27	0.0100

of about 25 data points. The weights were chosen in such a way that periods longer than about 50 units of time, or shorter than about 13 units, would be eliminated. The actual frequency response of this filter is shown in Figure 2. The ordinate of this curve gives

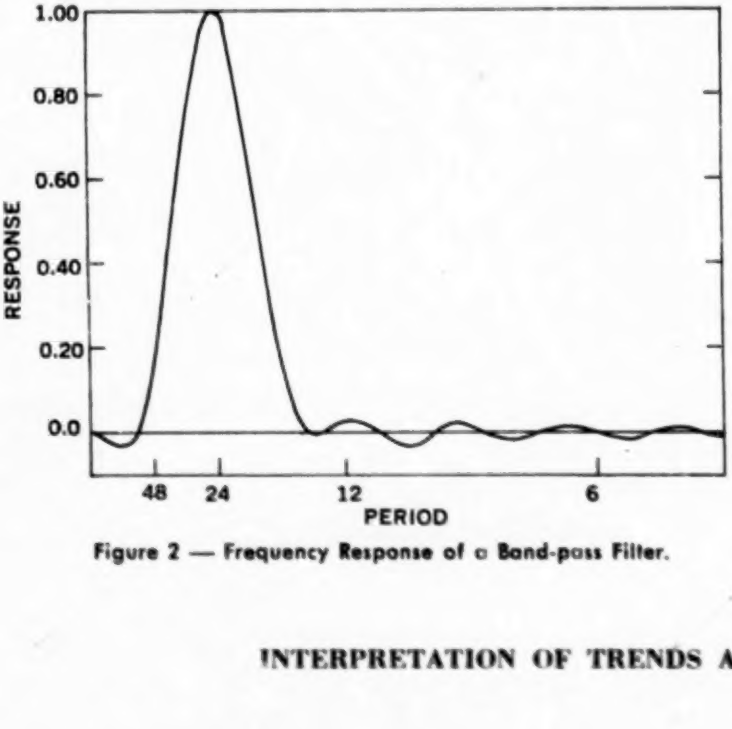


Figure 2 — Frequency Response of a Band-pass Filter.

the ratio of the amplitude of a wave of a given frequency f in the time series after filtering to the original amplitude of the wave before filtering. The frequency response R_f of a filter is a function of frequency and is given by the formula

$$R_f = W_0 + 2 \sum_{K=1}^{n} W_K \cos 2 \pi f K$$

where f is expressed in terms of cycles per data interval and ranges from 0 to 1/2.

If, on the other hand, R_f is specified, the weights W_K can be determined by the formula

$$W_K = R(0) + 2 \sum_{f=½n}^{n/2n} R(f) \cos 2 \pi f K,$$

$$f = 1/2n\ 2/2n, 3/2n, \ldots, 1/2$$

Further details of these procedures can be found in Brier.[3]

The filter shown in Table 1 has been applied to over 200 years of monthly precipitation data for England. A sample of the data is plotted in Figure 3; the corresponding filtered output F_t is shown, on an amplified scale, in Figure 4. The purpose of this study was to investigate whether there was any period of around 24 to 27 months that maintained a constant amplitude or phase over the 200 years. The results were negative or inconclusive. Another study was made to learn whether the peaks in Figure 4 (and the data for the remaining years) tended to be more frequent or to have greater amplitude during some calendar months than others. This was done by tabulating for each peak the amount of the deviation above the zero line and the month of occurrence. These were plotted in the polar diagram of Figure 5, which shows an essentially random distribution. Thus there is no strong evidence of any period of around 24 or 26 months that is phase-locked with the annual cycle.

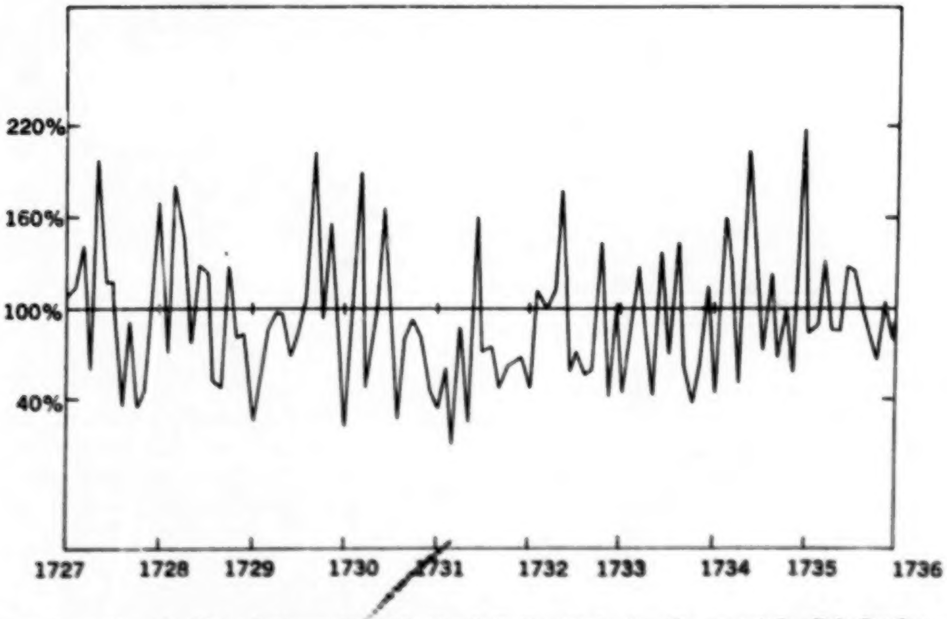


Figure 3 — Sample Plot of Mean Monthly Precipitation Data for Group of English Stations.

In the discussion on spectrum analysis it could have been pointed out that this method of analysis is appropriate when the contributions to the total variance result from a continuous spectrum of frequencies. If there are lines in the spectrum corresponding to genuine periodic terms, then it is usually considered desirable to remove their effects. The difficult problem may be to determine whether there is a line. If the amplitude of a true periodic component is small, the line may be hidden in the noise

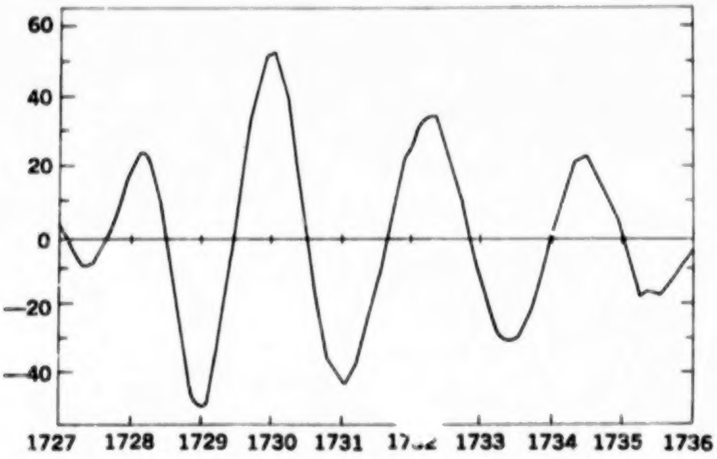


Figure 4 — Output of Filter Used on Monthly Precipitation Data for Group of English Stations.

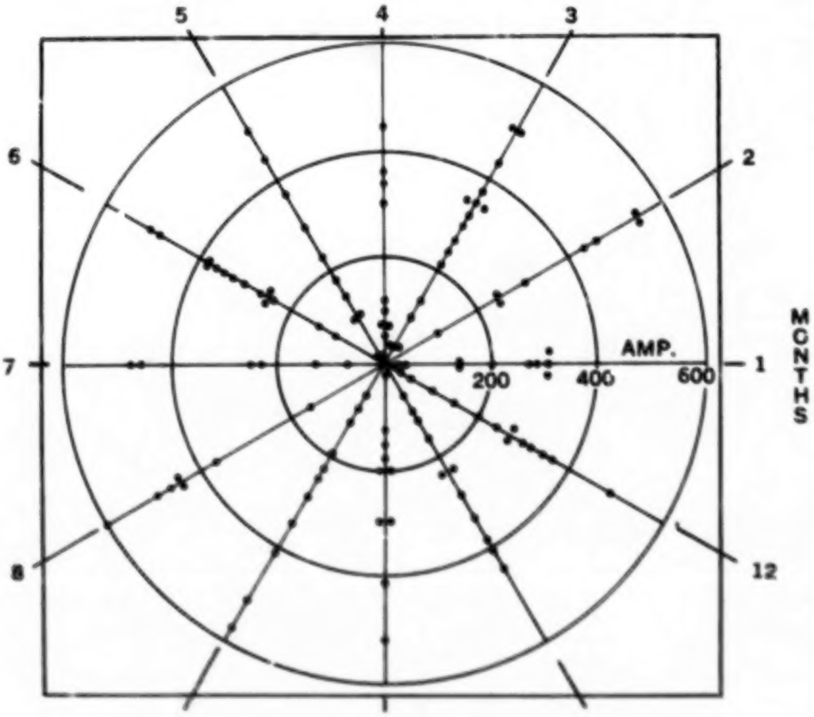


Figure 5 — Distribution of Amplitude and Phase of Peaks in Filtered Series of English Precipitation Data.

and is not likely to be detected by spectral analysis. This is not the place to discuss this problem in detail, but one suggestion can be made. If a true period of length p exists in the data, then it should persist through the entire time series without any significant change in phase. If the entire record is broken into two equal parts, a Buys-Ballot table can be constructed for the first half and second half of the record independently. If, for example, one is interested in examining the time series for a period of 27 days, the data are arranged in 27 columns with day 1, 28, 55, etc. placed in the first column. Days 2, 29, 56 . . . etc. are placed in the second column, and this procedure is followed until the data are all used and the averages for each column determined. Figure 6 shows the results of using this procedure for 50 years of precipitation data

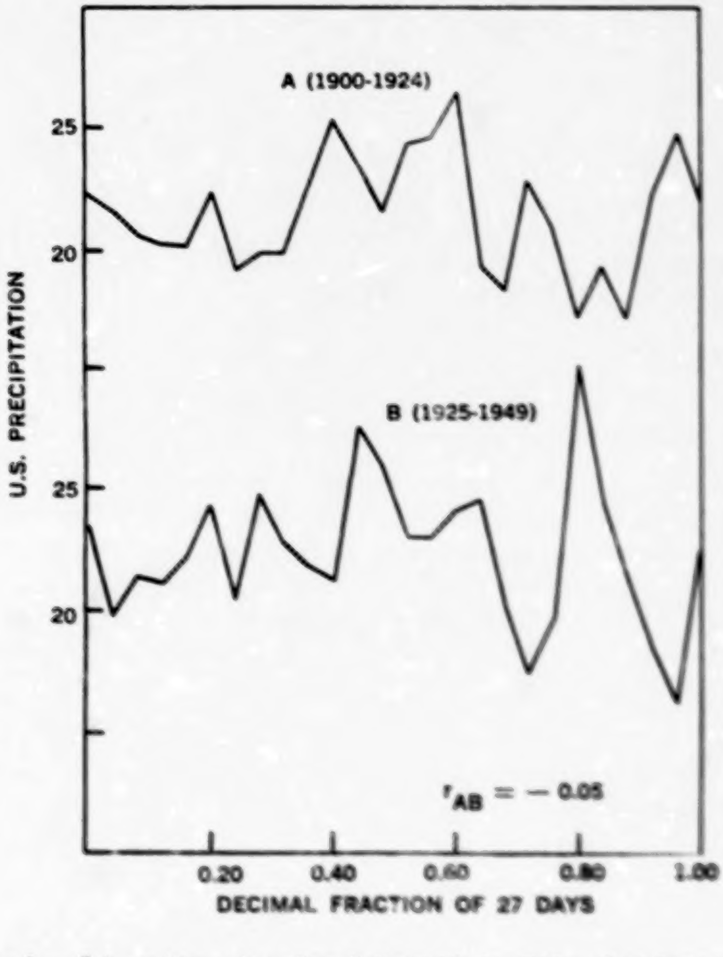


Figure 6 — Column Means from Buys-Ballot Table for 27-Day Trial Period. Curve A, U. S. Precipitation Data 1900-1924; Curve B, U. S. Precipitation Data 1925-1949.

for 1544 weather stations in the United States. Use of a high-speed electronic computer for computation makes it more convenient to plot the data in terms of the decimal fraction of the period being examined. In this diagram, there is little or no resemblance between the curves for the two independent periods, the correlation being $r_{AB} = -0.05$. With the computer it was practical to examine periods from 27.000 days to 31.000 days by intervals of 0.005 day. The highest correlation between A and B was found for a period of 29.530 days ($r_{AB} = 0.71$), which corresponds to the lunar synodic period. Figure 7 shows a partial plot of these results, which confirm the findings of Bradley et al.[3]

Although many additional operations can be applied to the analysis of time series, the main point I would like to make is that the modern high-speed computer enables one to do a great many things economically and in much greater detail than would have been considered possible (or even desirable) as little as 5 years ago.

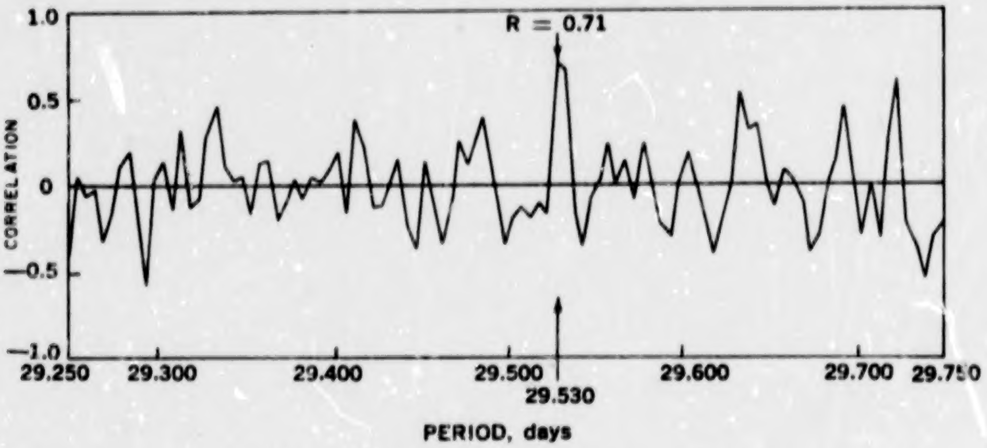


Figure 7 — Correlation Coefficients Between Column Means for Two Independent Time Periods for Trial Periods Extending from 29.250 Days to 29.750 Days.

REFERENCES

1. Blackman, R. B. and J. W. Tukey. 1958. The measurement of power spectra from the point of view of communications engineering. Bell System Tech. J. *37*: 185-282, 485-569.

2. Bradley, D. A., Woodbury, M. A., Brier, G. W. 1962. "Lunar Synodical Period and Widespread Precipitation." *Science* 137: 748-749.

3. Brier, G. W. 1961. "Some Statistical Aspects of Long-Term Fluctuations in Solar and Atmospheric Phenomena." Annals of the New York Academy of Sciences. 95: 173-187.

4. Holloway, J. L., Jr. 1958. Smoothing and filtering of time series and space fields. Advances in Geophysics. IV: 351-389. Academic Press, New York, N. Y.

5. Jenkins, G. M. 1961. "General Considerations in the Analysis of Spectra." Technometrics. 3: 133-190.

6. Mitchell, J. M., Jr. 1963. "Some Practical Considerations in the Analysis of Geophysical Time Series." (to be published).

7. Panofsky, H. A. and G. W. Brier. 1958. Some Application of Statistics to Meteorology. Pa. State Univ., University Park, Pa.

8. Tukey, J. W. 1949. The sampling theory of power spectrum estimates. Symposium on Applications of Autocorrelation Analysis to Physical Problems. Woods Hole, Mass. pp. 47-68.

9. Ward, F. W., Jr. 1960. The variance (power) spectra of C_i, K_p, and A_p. J. Geophys. Research. 65: 2359-2373.

10. Wiener, N. 1930. Generalized harmonic analysis. Acta Math. 55:117-258.

11. Wiener, N. 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Technology Press of M.I.T., Cambridge, Mass.

Dr. L. D. Zeidberg
and
Emanuel Landau
School of Medicine
Department of Preventive Medicine
and Public Health
Vanderbilt University, Nashville, Tennessee

SUMMARY

The hypotheses on which research is based must be limited in scope so that a measurable aspect of a problem can be defined with precision. Ultimately a chain of proved subsidiary hypotheses may serve to validate the major program objective, which is usually based on a major hypothesis.

In the Nashville Air Pollution Study the broad objective was to determine whether health is adversely affected by air pollution. It was postulated that health is affected and that the effects of air pollution are measurable. Four studies were designed; two of these are described in detail to show how various hypotheses were developed and tested, what conclusions were drawn, and what further avenues of research were opened as a result of these studies.

DATA INTERPRETATION — DRAWING CONCLUSIONS

INTRODUCTION

Research generally is based on the development and the testing of hypotheses. This is not only the scientific approach, but is also part of the accepted epidemiologic method. Hypotheses usually have some basis in already established facts. The assembling of such facts is a necessary prelude to the development of hypotheses. Once developed, they must be subjected to searching tests made with scientific objectivity. Testing generally involves the collection and interpretation of data, from which conclusions may be drawn that will either validate the hypotheses or negate them.

In order to define a measurable aspect of a problem with precision, hypotheses must of necessity be limited in scope. It may be necessary to develop and test a whole chain of subsidiary but related hypotheses in order to marshall the data required to validate the major program objective. In all of this, data interpretation is a key step that leads ultimately to conclusions. It is well to recall the often-quoted words of Frost: "Epidemiology at any given time is something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bounds of direct observation."[1]

Two of the air pollution studies[2, 3] conducted in Nashville, Tennessee, under a contract with the Air Pollution Division of the Public Health Service[4] will be examined to illustrate how the results of the analyzed data were related to initial hypotheses, what conclusions were drawn, and what new hypotheses were formulated.

PULMONARY ANTHRACOSIS AS AN INDEX OF AIR POLLUTION

Pulmonary anthracosis is a condition of the lungs in which black pigment is deposited as a result of the inhalation of particles of coal dust, and perhaps of other dusts. Pathologists in Nashville had gained an impression, not tested, however, by definitive studies, that Nashville residents had more such pigment in their lungs than did non-residents. If this were true, it should be possible to use the pigmentation of the lung as an index of air pollution due to combustion of coal in a community. Several hypotheses were advanced, and a plan was devised to test them. The hypotheses were:

1. Anthracosis in the lungs of Nashville residents is directly related to air pollution in Nashville.

2. The degree of anthrocosis will vary among Nashville city and out-of-city residents.

3. Among Nashville residents the degree of anthracosis will vary depending on the length of residence in the city.

4. Occupational exposure to coal dust may be a factor that affects the degree of anthracosis, but not to the exclusion of other exposure.

5. Anthracosis is a cause of ill health.
   (a) Anthracosis is associated with specific symptoms; and
   (b) Anthracosis is related to the occurrence of cardiorespiratory disease.

To test these hypotheses a series of consecutive autopsies (except those in subjects under 5 years of age) done at Vanderbilt University Hospital between 1953 and 1956 was studied. The degree of anthracosis in the lungs of 641 subjects was evaluated according to standards established by Dr. John Shapiro, professor of Pathology at Vanderbilt University School of Medicine. The lungs were classified as showing no pigment, or showing minimal, moderate, or severe anthracosis. The residence of each subject at the time of death was the determining factor in designating him as a Nashville or out-of-city resident. For persons in the first category, city directories were searched at 5-year intervals to establish how long they had lived in the city, and in what part of it. The residential data were then converted into a classification of exposure to low, moderate, or high pollution, on the basis of aerometric data collected in the engineering phase of the study.[5] For those who were out-of-city residents at the time of death, it was assumed that they had never lived in Nashville. For city residents occupational data also were sought by reference to city directories. The data concerning symptomatology and pathology were obtained from the hospital records and the autopsy protocols, respectively.

In development of the plan of the study it was necessary to make certain assumptions, the validity of which may be questioned. It was assumed, for example, that out-of-city residents had always lived out of the city, as noted above. In comparison of Nashville and non-Nashville residents it was assumed that out-of-city dwellers did not live in a city and were not exposed to as much coal smoke as residents of Nashville. This assumption would tend to minimize differences in anthracosis in the two groups, and therefore would make real differences even more significant.

Another major assumption was that the level of air pollution in any specific area of the city in the period from 1958 to 1959 was indicative of that prevailing as long as 20 years ago. Although abundant evidence indicated that the pollution situation in Nashville had improved considerably with the years, it was assumed that the improvement had been proportionately comparable in different parts of the city.

The bias inherent in autopsy material would ordinarily make it extremely hazardous to extrapolate or apply such data to a general population. It was assumed, however, that anthracosis would not necessarily bring people to Vanderbilt University Hospital, or prove fatal, or predispose them to post-mortem examination.

In testing of the hypotheses that had been formulated at the outset, it was important to exercise extreme care to avoid the introduction of bias. Consequently this was conducted as a "blind" study, with each phase entrusted to a different investigator, who worked independently, so that the one responsible for classifying the degree of anthracosis in the lungs of subjects, for example, identified them only by number and had no knowledge of their age, sex, race, residence, or occupation. The importance of such safeguards cannot be emphasized enough. Where absolute objectivity is essential, as it must be in any scientific investigation, even the slightest bias could destroy the validity of interpretations and conclusions.

In testing of the first hypothesis, that anthracosis is directly related to the level of air pollution in Nashville, multiple regression techniques[6] were used. The relationship between residential exposure to different levels of air pollution and the amount of anthracosis in the lungs was measured, with age considered as a variable. Analyses were done separately for males and females. For males, the multiple correlation coefficient was 0.489, indicating a highly significant relationship. While the individual regression coefficients for residence alone, not considering age, were not significantly different from zero, there was a direct relationship between degree of anthracosis and degree of exposure to air pollutants. For females, the multiple correlation coefficient and the individual regression coefficients for residence were highly significant. One can speculate that the female, who is apt to be more closely related to the residential environment than is the male, reflects a reaction to the home environment, uninfluenced by an occupational exposure. It may be concluded, therefore, that the first hypothesis may have reasonable validity.

The second hypothesis, that the degree of anthracosis varies in Nashville and out-of-city residents, was tested by comparing the degree of anthracosis for both groups, by age. These comparisons are shown in Figure 1. Above the age of 25, Nashville residents showed a consistently higher level of severe anthracosis. Thus, the second hypothesis appears to be valid also.

For the third hypothesis, that in Nashville residents the degree of anthracosis varies with the length of residence in the city, the analysis was limited to the 466 subjects who were over 45 years of age, because age could be a limiting factor. The subjects were divided into two groups: those with residence less than 20 years, and those with residence of 20 years or more. Figure 2 shows this comparison and indicates that the longer-term residents had more anthracosis. Very few of those who had lived in the city for more than 20 years showed minimal anthracosis. Figure 2 also shows the comparative degree of anthracosis among out-of-city residents, but this merely another way of illustrating the data in Figure 1. The third hypothesis may be considered valid also.

The fourth hypothesis suggested that occupational exposure to coal dust affecting the degree of pulmonary anthracosis would not exclude the effect of other factors, such as the residential environment. It was not possible to put this hypothesis to the test because of the insufficiency of occupational data. The city directories that were searched for occupational listing of the 329 Nashville residents in the autopsy group were unexpectedly deficient in this regard. For 129 females, many of them housewives, data were totally lacking. An additional 17 were of the younger age group, and unlikely to have been employed. The information obtained was believed to be too meager for interpretation. Here is an illustration of the limitations of retrospective studies, which are perforce limited qualitatively and quantitatively to already recorded data in the absence of additional followup activity.

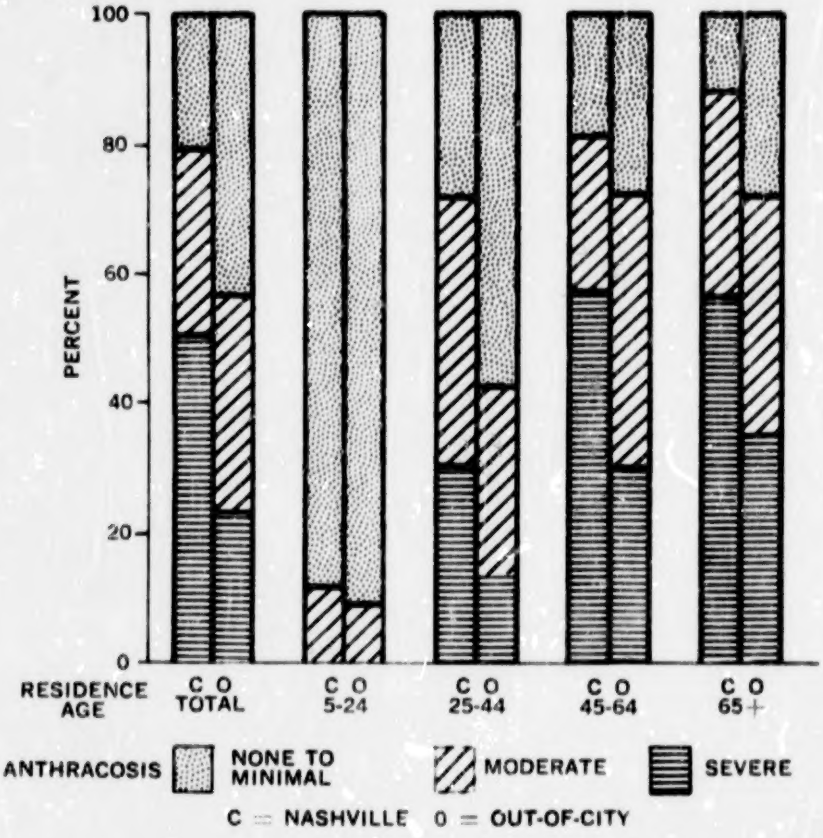


Figure 1 — Age and Residence Differences in the Degree of Anthracosis Found at Autopsy in 641 Individuals, Vanderbilt University Hospital, 1953-56.

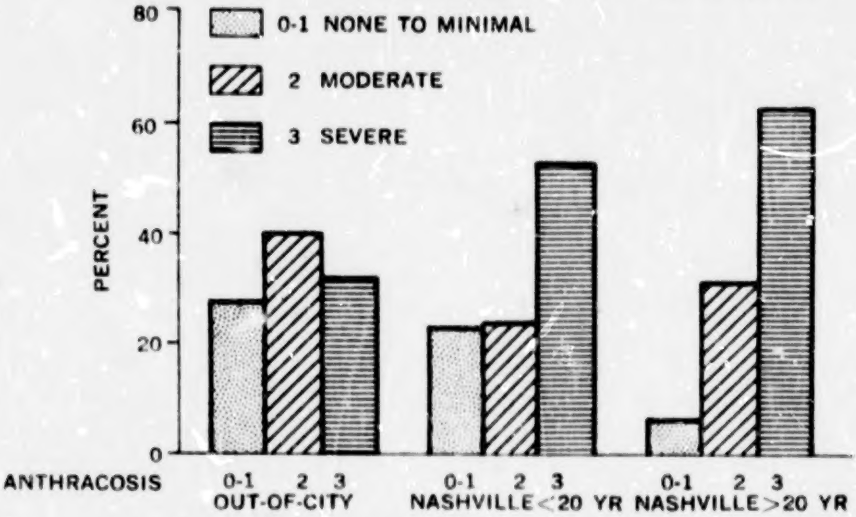


Figure 2 — Residence Differences in the Degree of Anthracosis Found at Autopsy in 466 Individuals 45 Years of Age and Over, Vanderbilt University Hospital, 1953-56.

The final hypothesis, related to anthrocosis as a cause of ill health, was formulated in two parts. The first of these was that anthracosis is associated with specific symptoms. To test this subhypothesis, the hospital records of each subject were studied carefully for symptomatology relating specifically to the cardiorespiratory system. Only the 560 white subjects in the autopsy group were included in this analysis. It was concluded that anthracosis was not characterized by specific cardiorespiratory symptoms. The second part of the hypothesis suggested that anthracosis is related to the occurrence of cardiorespiratory disease. The data to test this hypothesis were obtained from hospital records and autopsy protocols. No specific disease could be related to anthracosis. One of two possible conclusions may be reached: either there is in fact no related symptomatology or pathology, or the data obtained are not sufficently accurate. The first of these conclusions has support among some pathologists[7, 8] but not among others.[9, 10] It is possible that our data are faulty and do not reveal pulmonary disease that was actually present. British investigators report the presence of a focal emphysema, or dilatation of air spaces in the lungs, associated with anthracosis.[11-13] In their studies they inflated the collapsed lungs to their normal size at autopsy. Our studies are based on small sections of collapsed lung, in which it would not be possible to observe the changes described by the British. A recent study in this country, however, refutes the British work and claims that focal emphysema is the cause, rather than the result, of deposits of anthracotic pigment in the lungs.[14] There the matter stands, and we may conclude that we have not been able to put this last hypothesis to an adequate test. To obtain an answer it will be necessary to set up new hypotheses and develop studies to test them. For example, it might be postulated that anthracosis follows rather than precedes the development of centrilobular emphysema. Retrospective studies could be planned if inflated lung specimens were available. If not, prospective studies would have to be done, and these might require a considerable period of time.

From our study of pulmonary anthracosis we may conclude that anthracosis in the lungs of Nashville residents is a fairly good index of the degree of air pollution to which they have been exposed during their residence, but we are unable to show that such deposits were necessarily injurious to their health.

MORBIDITY IN RELATION TO AIR POLLUTION

As further illustration of how research data may be interpreted and conclusions drawn, some of the features of a morbidity survey conducted in Nashville will be discussed. The survey was part of a general study of the health effects of air pollution. The following hypotheses were formulated:

1. The morbidity experience of Nashville residents is directly related to the level of air pollution in their environment.

2. Illness due to specific causes, particularly respiratory and cardiovascular, will vary according to the levels of air pollution in different areas.

3. Specific age groups will be affected differently.

4. Occupational exposure will affect the occurrence of illness, but not to the exclusion of other exposures.

To test these hypotheses a survey of a representative sample of the population was planned, by means of direct interview of a responsible adult in each of the selected households. Morbidity in the middle socio-economic class was analyzed because members of this class were found in all levels of pollution. Further, they comprised the largest group in the surveyed population.

The testing of the hypotheses did not produce general validation. For the first hypothesis, that morbidity is directly related to air pollution exposure, no regular pattern could be shown for any of the four pollutants studied. For white residents over 55 years of age, however, in whom the effects of prolonged exposure to air pollution might be expectedly most pronounced, a consistent pattern of increasing morbidity with increasing exposure to air pollutants was observed when the soiling index and 24-hour SO_2 concentration were used as indexes (Figure 3). Because too few of the non-whites lived in areas of low pollution, comparisons were made between residents of high and moderate pollution areas only. These comparisons showed the same patterns for non-white females as observed for the white residents, but for the non-white males only the soiling index showed a significant correlation (Figure 4).

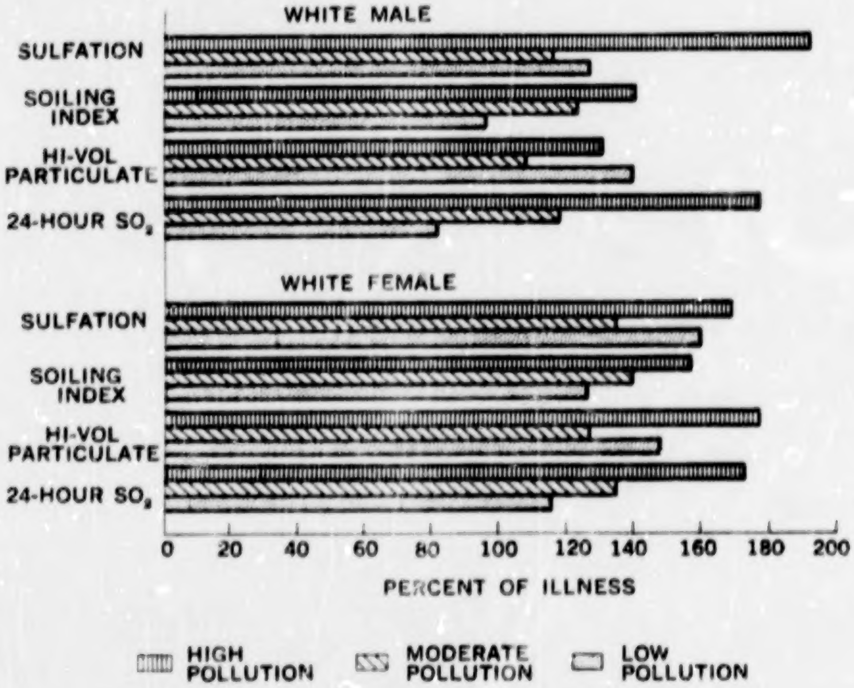


Figure 3 — Percent of Illness for All Causes During the Year Prior to the Survey Among White Middle Class Individuals 55 Years of Age and Over, by Sex and by Degree of Exposure to Atmospheric Pollutants. Nashville Air Pollution Study.

For the second hypothesis, that morbidity for specific causes (such as respiratory and cardiovascular disease) is directly related to air pollution exposure, partial validation could be shown. No correlation could be shown for respiratory illness, cancer, or gastrointestinal disease; cardiovascular morbidity increased with exposure to particulates measured by the soiling index and to SO_2 for white males, while for white females direct associations with all four pollutants were observed (Figure 5). Among non-whites no consistent pattern could be shown.

For the third hypothesis, that specific age groups will be affected differently by exposure to air pollutants, strong support was evidenced. Only for those over 55 years of age could any pattern of relationship between morbidity and air pollution be shown. At this point a new hypothesis might be advanced, that the effects of usual exposure to air pollution become manifest only after prolonged experience.

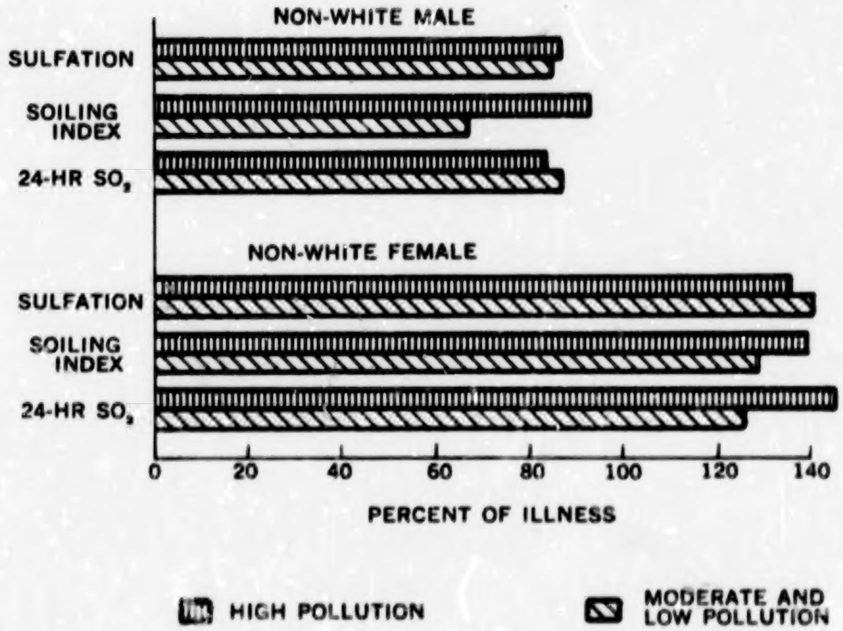


Figure 4 — Percent of Illness for all Causes During the Year Prior to the Survey Among Non-White Middle Class Individuals 55 Years of Age and Over, by Sex and by Degree of Exposure to Atmospheric Pollutants. Nashville Air Pollution Study.

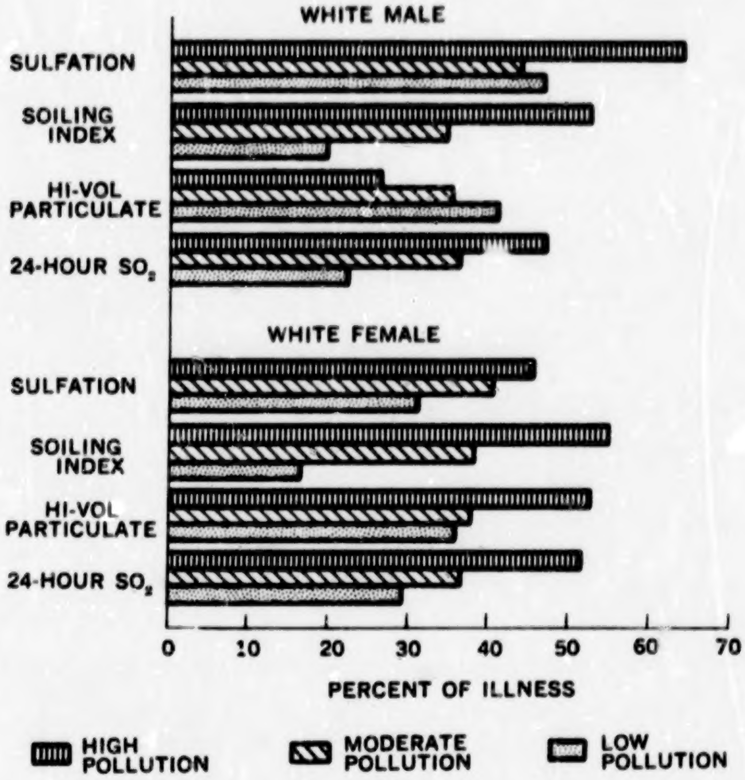


Figure 5 — Percent of Cardiovascular Illness During the Year Prior to the Survey Among White Middle Class Individuals 55 Years of Age and Over, by Sex and by Degree of Exposure to Atmospheric Pollutants. Nashville Air Pollution Study.

The fourth hypothesis, that occupational exposure will affect the occurrence of illness, but not to the exclusion of other causes, may be advanced to explain the lack of correlation noted above, particularly for non-whites. In comparisons based on pollution levels in the residential environment, no account is taken of the influence of occupational environments. For workers, almost one-third of their exposure experience occurs away from home. This may account for the lack of correlation between morbidity and air pollution exposure. In order to refine the data, an analysis was made of the morbidity experience of females 15 to 64 years of age, classified into two groups, working and housekeeping. The latter would be expected to reflect more accurately the influence of the residential environment alone. Because subdivision of the data by socio-economic class and by specific cause produced cells too small for analysis, only total morbidity was considered, and rates were adjusted for age. Morbidity rates were higher in general for the white housekeeping females than for the working females. The former groups showed a direct relationship between morbidity and level of pollution for all pollutants except soiling index. For this pollution index the morbidity rates were highest in the high pollution areas, but no difference could be shown in morbidity rates for moderate and low exposure (Figure 6). For the white working females, none of the pollutants showed any correlation with morbidity experience. For the non-whites, those few who lived in areas of low pollution were combined with the moderate pollution group for comparison with those in high pollution areas. Both the working and housekeeping non-white females showed similar patterns wherever numbers were large enough to allow age

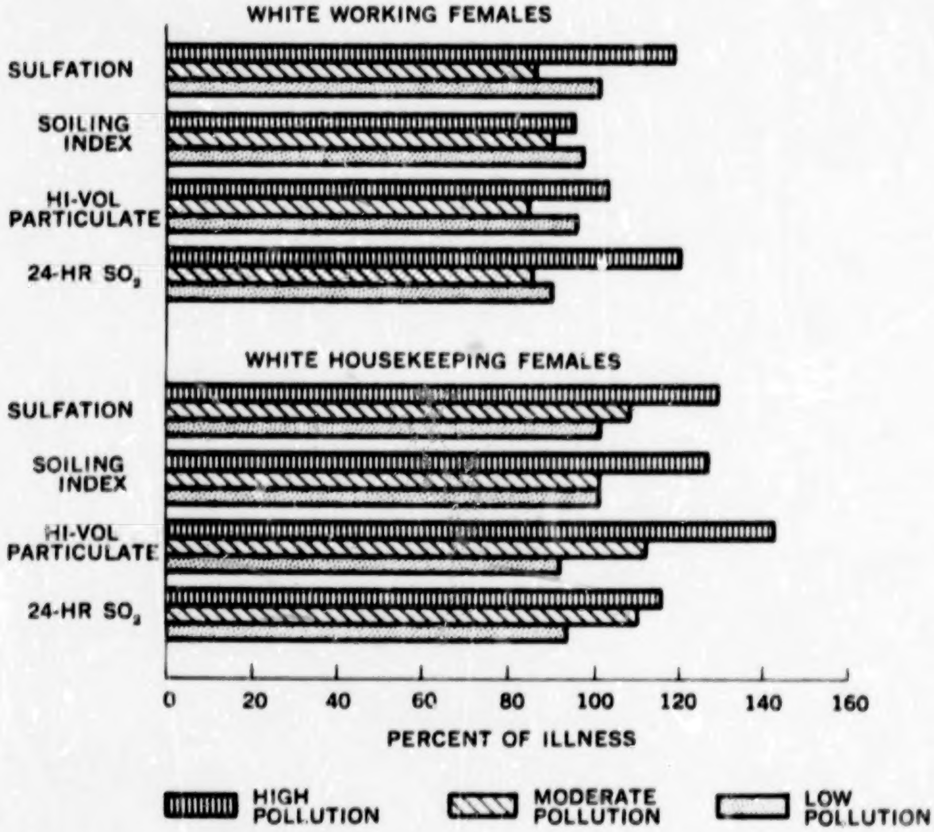


Figure 6 — Age-Adjusted Morbidity Rate for All Causes for the Year Prior to the Survey Among White Working and Housekeeping Females 15-64 Years of Age, by Exposure to Atmospheric Pollutants. Nashville Air Pollution Survey.

adjustment (Figure 7). Since the non-white female is often employed as a domestic, her occupational environment is not likely to be in the commercial and industrial areas of the city where pollution is greatest. Since her occupational environment may not influence her experience significantly, there may be no difference between her morbidity experience and that of the housekeeping non-white.

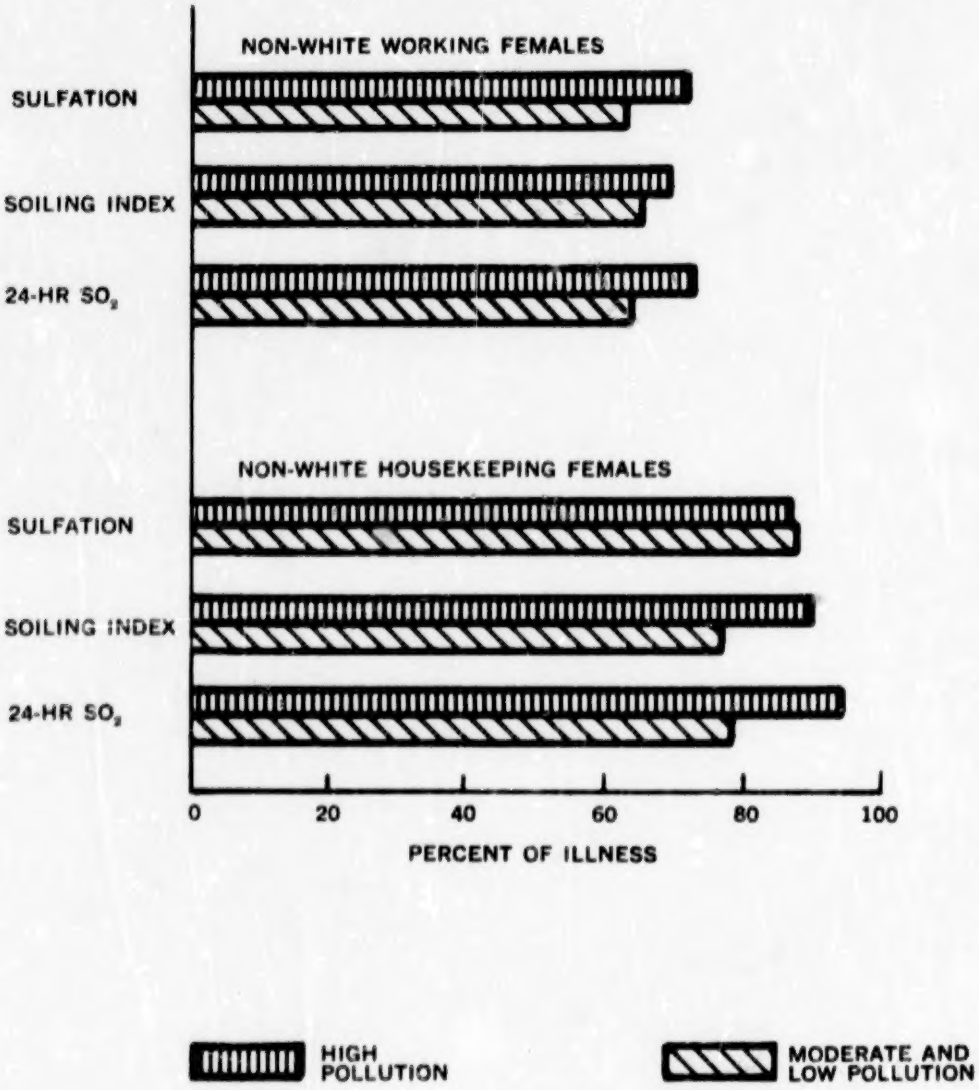


Figure 7 — Age-Adjusted Morbidity Rate for all Causes for the Year Prior to the Survey Among Non-White Working and Housekeeping Females 15-64 Years of Age, by Exposure to Atmospheric Pollutants. Nashville Air Pollution Survey.

In interpreting the results of an investigation great care must be taken to discover hidden bias or error that may seriously influence the data. For example, it was well known from other studies[15-18] and from our own experience in a pre-test of the questionnaire used in these studies that the interview method does not provide accurate qualitative or quantitative measures of the prevalence of illness. Since we were primarily concerned with the relative frequency of illness in areas with differing levels of air pollution however, it was assumed that biases or errors of response would not vary with differences in exposure to air pollutants. Analysis of the data showed that this assumption was grossly invalid. For, although there is nothing about exposure to high or low levels of pollution that should produce differences in awareness or reporting of illness, there was a marked relationship between socio-economic status and exposure to pollutants in the residential environment. In general, the lower socio-economic classes lived in the most polluted areas. The socio-economic influence on such factors as education, occupation, and utilization of medical services is too well known to require elaboration. These factors could seriously affect the reporting of illness, and may explain the unexpected finding of more illness reported in upper-class children under 15 years of age than in middle-class children, and even more than in the lower-class children. To minimize or eliminate possible bias, comparisons of morbidity in relation to air pollution should be made by holding the socio-economic factor constant.

The final conclusion reached by the interpretation of the morbidity study data is that the hypotheses advanced have been, at best, only partially validated. This is rather inconclusive and may mean either that there is no true relationship between exposure to air pollutants and illness or that the morbidity data obtained are not sufficiently reliable. It is difficult to accept the first of the two possibilities in the light of what is known about the effects of exposure to high concentrations of pollutants such as has been reported in London[19] and in Donora, Pennsylvania.[20] To the authors, the second possibility seems to be more plausible. This means that more accurate methods for measuring morbidity in a community must be employed in studies of this kind. It should be observed that no note of suspicion concerning the reliability of the aerometric measurements has been sounded.

The authors have drawn on two of the four phases of the Nashville Air Pollution Study to illustrate how hypotheses are established and tested, and how new hypotheses are then promulgated for further testing.

REFERENCES

1. Frost, Wade Hampton. In the introduction to *Snow on Cholera,* New York, The Commonwealth Fund, 1936.

2. Zeidberg, L. D., and Prindle, R. A. The Nashville Air Pollution Study: II. Pulmonary Anthracosis as an Index of Air Pollution. Am. J. Public Health, *53*:185-199, Feb. 1963.

3. Zeidberg, L. D., Prindle, R. A., and Landau, E. The Nashville Air Pollution Study: III. Morbidity in Relation to Air Pollution. Presented before the Epidemiology Section, at the 90th Annual Meeting of the American Public Health Association in Miami Beach, Florida, October 17, 1962. Accepted for publication in the American Journal of Public Health and scheduled to appear in the December, 1963, issue.

4. U. S. Department of Health, Education, and Welfare, Public Health Service, Air Pollution Medical Program Contracts No. SAph 68348 and 69628.

5. Zeidberg, L. D., Schueneman, J. J., Humphrey, P. A., and Prindle, R. A. Air Pollution and Health: General Description of a Study in Nashville, Tennessee. J. Air Poll. Control Assoc., *11*:289-297, June, 1961.

6. Snedecor, G. W. Statistical Methods. Iowa State College Press, 1956.

GPO 814-105-10

7. Anderson, W. A. D. Pathology (3rd edition). St. Louis, Missouri, Mosby, 1957. p. 650.

8. Boyd, W. Textbook of Pathology (6th edition). Philadelphia, Pa., Lea and Febiger, 1953. p. 393.

9. Saphir, O. (Editor). A Textbook of Systemic Pathology. New York and London, Grune and Stratton, 1958, Vol. I, pp. 323-4.

10. Moore, R. A. A Textbook of Pathology (2nd edition). Philadelphia, Pa., Saunders, 1951, pp. 512-3.

11. Heppelston, A. G. Essential Lesion of Pneumokoniosis in Welsh Coal Miners. J. Path. & Bact., *59*:453-460, July, 1947.

12. Gough, J., James, W. R. L., and Wentworth, J. E. Comparison of Radiological and Pathological Changes in Coalworkers Pneumoconiosis. J. Fac. Radiologists. 1, 1:28-29, July, 1949.

13. Oderr, C. P. Emphysema, Soot and Pulmonary Circulation-Macroscopic Studies of Aging Lungs. JAMA, *172*:1991-1998, April 30, 1960.

14. Pratt, P. C., Jutabha, P., and Klugh, G. A. The Relationship Between Pigment Deposits and Lesions in Normal and Centrilobular Emphysematous Lungs. Am. Rev. Resp. Dis., *87*:245-256. February, 1963.

15. Cobb, S., Thompson, D. J., Rosenbaum, J., Warren, J. E., and Merchant, W. R. On the Measurement of Prevalence of Arthritis and Rheumatism from Interview Data. J. Dis. of Children, *3*:134-139, Feb., 1956.

16. Trussell, R. E., Elinson, J., and Levin, M. Comparison of Various Methods of Estimating the Prevalence of Chronic Disease in a Community — the Hunterdon County Study. Am. J. Pub. Health, *46*:173-182, Feb., 1956.

17. Krueger, D. E. Measurement of Prevalence of Chronic Disease by Household Interviews and Clinical Evaluations. Am. J. Pub. Health, *47*:953-960, Aug., 1957.

18. Lilienfeld, A. M., and Graham, S. Validity of Determining Circumcision Status by Questionnaire as Related to Epidemiological Studies of Cancer of the Cervix. J. Nat. Cancer Inst., *21*:713-720, Oct., 1958.

19. Committee on Air Pollution: Interim Report, Her Majesty's Printing Office, London, 1956.

20. Schrenk, H. H., Heimann, H., Clayton, G. D., Gafafer, W. M., and Wexler, H. Air Pollution in Donora, Epidemiology of the Unusual Smog Episode of October, 1948: Preliminary Report. Public Health Bulletin No. 306, 1949.

DISCUSSION: INTERPRETATIONS AND CONCLUSIONS

PANEL MEMBERS

G. J. Taylor
Assistant Chief, Bureau of Air Sanitation
California Dept. of Health, Berkeley

Robert A. McCormick
Chief, Meteorological Section, DAP
U. S. Public Health Service, Cincinnati

Dr. R. O. McCaldin
Deputy Chief, Field Studies Branch, DAP
U. S. Public Health Service, Cincinnati

Dr. Harold J. Paulus
Associate Professor of Public Health
University of Minnesota, Minneapolis

Mr. McCormick noted the necessity for extreme caution in research, the conscious thinking through of any problem before beginning work, and the importance of controls and auxiliary tests to reduce the number of possible interpretations of experimental results. He quoted the statement, attributed to Darwin, that "nature will tell you a direct lie if she can."

The group was asked for a prediction as to when, if ever, air quality measurements would be as complete and representative as meteorological measurements are presently. Mr. McCormick replied that meteorological data are not as representative as commonly thought and that the CAMP stations provide data as representative as meteorological data, although at a cost tremendously higher than wind and temperature instrument installations. He noted that the Weather Bureaus receive a large number of inquiries from the public, and predicted better air quality data when public interest in air pollution becomes comparable.

Mr. Taylor commented that the amount of air monitoring usually is proportional to the magnitude of the problem and the level of public recognition. At present, needs for air pollution data are not similar throughout the nation, as are the requirements for meteorological measurement; nor are air pollution situations usually the same over vast areas, as are meteorological phenomena. Air quality measurements in the near future will still be confined to problem areas, although offering much better coverage of these areas. Mr. Linsky disagreed with Mr. Taylor and observed that the requirements for meteorological measurements are as broad and varied as are those for air pollution data. He expressed concern over the seeming lack of interest in interferences that prevent the measurement of what is intended to be measured, particualrly in air quality measurements, and cautioned against blind acceptance of routine data.

Mr. Bryan asked how long air quality measurements must be made to permit time series analyses as discussed by Mr. Brier. Mr. McCormick replied that formulas are available that describe precisely the data required to yield a given precision in series analysis. Dr. McCaldin noted that economics usually dictate in the air pollution field: the decision generally is based on how much can we afford, rather than how much we need.

Mr. Gruber asked what natural geophysical cycles influence air quality, and how long we must sample to include the effect of the most important of these climatological cycles. Mr. Bellamy pointed out that on the power spectra displayed by Mr. Brier the longest cycles were the largest, and that this would seem to indicate the necessity for continuous sampling on a permanent basis. He noted that continuous detailed data will become more readily available when air pollution control agencies use them in their operations in the same manner that airports presently use meteorological data. Mr. Schueneman commented that probability techniques similiar to those used by hydrologists might be useful when records of air quality over long periods become available.

The participants were asked how far one can deviate from an ideal sampling situation in the interest of practicality without compromising the data obtained. Mr. Taylor observed that water, sewerage, electric power, land, and economic requirements severely limit sampling sites, particularly for large installations such as CAMP stations. He pointed out, however, that present aid quality monitoring is generally not aimed at research, but at assessing the air pollution situation and getting information to form research hypotheses, and hence concern should be directed not only at deviations from ideal sampling but also at deviations from program objectives. Mr. Gruber pointed out that different contaminants demand different criteria in selection of sampling locations; for example, dustfall is more location-dependent than soiling index because of the larger, rapidly settling particles involved. Mr. Schueneman suggested screening various locations with simple portable samplers to determine location-dependence before selecting sites. He also pointed out that one must consider limitations in site selection, sampling, and analysis when determining the degree of confidence in the results. Mr. Litsky added that sampler location can often be guided by program objectives, i.e., characterization of over-all air quality, evaluation of large single sources, or consideration of many small sources. He reminded the group that although many air pollution "sensors" measure only an effect of air pollution rather than physical or chemical quantities, these measurements are valid and should be encouraged.

Dr. Zeidberg mentioned that in the PHS Nashville study an attempt was made to determine the minimum sampling necessary to characterize the air, but the results may not be applicable beyond the Nashville area. He warned against injudicious drawing on the experience of others in such matters.

SESSION 8: Measurements of Water Environment

Chairman: **Maurice LeBosquet**
Office of the Chief
Division of Water Supply and Pollution Control
U. S. Public Health Service

Dr. H. B. N. Hynes*
Department of Zoology
University of Liverpool, England

SUMMARY

The complexities of biological reactions to water conditions are reviewed, as well as the problems in presentation of biological results in a form that can be readily understood by workers in other disciplines. The problems of biological sampling are also considered.

It is concluded that no system of presentation of results so far proposed is really satisfactory, and any system that does not include tabulated raw data is concealing information that should be recorded.

Interpretation of biological data requires considerable training and must be left to biologists, who also must design the sampling program for each particular situation. Biology, like medicine, is too complex a subject to codify, but it is an essential tool in the full evaluation of water quality.

THE INTERPRETATION OF BIOLOGICAL DATA WITH REFERENCE TO WATER QUALITY

Sanitary engineers like to have data presented to them in a readily assimilable form and some of them seem a little impatient with biologists who appear unable to provide definite quantitative criteria applicable to all kinds of water conditions. I think the feeling tends to be that this is the fault of biologists, and if they would only pull themselves out of the scientific stone-age all would be well. I will try to explain here why I believe that biological data can never be absolute nor interpretable without a certain amount of expertise. In this respect biologists resemble medical men who make their diagnoses against a complex background of detailed knowledge. Anyone can diagnose an open wound but it takes a doctor to identify an obscure disease; and although he can explain how he does it he cannot pass on his knowledge in that one explanation. Similarly, one does not need an expert to recognize gross organic pollution, but only a biologist can interpret more subtle biological conditions in a water body; and here again he can explain how he does it, but that does not make his hearer a biologist. Beck (1957) said something similar at a previous symposium in this city in 1956.

THE COMPLEXITY OF BIOLOGICAL REACTIONS TO WATER CONDITIONS

The aquatic habitat is complex and consists not only of water but of the substrata beneath it, which may be only indirectly influenced by the quality of the water. Moreover, in biological terms, water quality includes such features as rate of flow and temperature regime, which are not considered of direct importance by the chemist. To many animals and plants maximum summer temperature or maximum rate of flow is just as important as minimum oxygen tension. The result is that inland waters provide an enormous array of different combinations of conditions, each of which has its own community of plants and animals; and the variety of species involved is very great. Thus, for example, Germany has about 6000 species of aquatic animals (Illies 1961a) and probably at least as many species of plants. Yet Europe has a rather restricted fauna because of the Pleistocene ice age; in most other parts of the world the flora and fauna are even richer.

---

* Now Chairman, Department of Biology, University of Waterloo, Ontario, Canada.

We know something about the way in which species are distributed in the various habitats, especially in the relatively much studied continent of Europe, but we have, as yet, little idea as to what factors or combination of factors actually control the individual species. Thus, it is possible to list the groups of organisms that occur in swift stony upland rivers (rhithron in the sense of Illies, 1961b) and to contrast them with those of the lower sluggish reaches (potamon). Similarly we know, more or less, the different floras and faunas we can expect in infertile (oligotrophic) and fertile (eutrophic) lakes. We are, however, much less informed as to just what ecological factors cause these differences. We know they include temperature and its yearly amplitude; oxygen, particularly at minimal levels; plant nutrients, such as nitrate, phosphate, silica, and bicarbonate; other ions in solution, including calcium, chloride, and possibly hydrogen; dissolved organic matter, which is necessary for some bacteria and fungi and probably for some algae; the nature of the substratum; and current. We also know these factors can interact in a complex manner and that their action on any particular organism can be indirect through other members of the biota. Thus, for example, heavy growths of encrusting algae induced by large amounts of plant nutrients, or of bacteria induced by ample supplies of organic matter, can eliminate or decimate populations of lithophile insects by simple mechanical interference. But the change does not stop there: the growths themselves provide habitats for the animals, such as Chironomidae and Naidid worms, which could not otherwise live on the stones. Similarly, if oxygen conditions over a muddy bottom reach levels just low enough to be intolerable to leeches, Tubificid worms, which the leeches normally hold in check, are able to build up to enormous numbers especially as some of their competitors (e.g. *Chironomus*) are also eliminated. One then finds the typical outburst of sludge worms, so often cited as indicators of pollution. This does not happen if the same oxygen tension occurs over sand or rock, however, as these are not suitable substrata for the worms. Many such examples could be given, but they would only be ones we understand; there must be a far greater number about which we know nothing. One must conclude, therefore, that quite simple chemical changes can produce far-reaching biological effects; that we only understand a small proportion of them; and that they are not always the same.

This seems like a note of despair, however, if water quality deviates too far from normal, the effects are immediately apparent. Thus, poisonous substances eliminate many species and may leave no animals (Hynes 1960); excessive quantities of salt remove all leeches, amphipods, and most insects and leave a fauna consisting largely of Chironomidae, caddis worms, and oligochaetes (Albrecht 1954); and excessive amounts of dissolved organic matter give rise to carpets of sewage fungus, which never occur naturally. Here no great biological expertise is needed, and there is little difficulty in the communication of results. It is when effects are slighter and more subtle that biological findings become difficult to transmit intelligibly to other disciplines.

THE PROBLEMS IN PRESENTATION OF BIOLOGICAL RESULTS

Because of these difficulties various attempts have been made to simplify the presentation of biological findings, but to my mind none of them is very successful because of the complexity of the subject. Early attempts at systematization developed almost independently on the two sides of the Atlantic, although they had some similarities.

In America, there was a simple division into zones of pollution, e.g. degradation, septic, and recovery, which were characterized in broad general terms. This simple, text-book approach is summarized by Whipple et al. (1947), and serves fairly well for categorizing gross organic pollution such as has been mentioned above. It was, however, soon found by Richardson (1929) during his classical studies on the Illinois River that typical "indicators" of foul conditions, such as Tubificidae and *Chironomus*, were not always present where they would be expected to occur. This was an early indication that it is not the water quality itself that provides suitable conditions for "pollution faunas," but other, usually associated, conditions — in this instance deposits of rich organic mud. Such conditions may, in fact, be present in places where water quality in no way resembles pollution, e.g., upstream of weirs in trout streams where autumn leaves accumulate and decay and cause the development of biota typical of organically polluted water. Samples must therefore be judged against a background of biological knowledge. Richardson was fully aware of this and was in no doubt about the condition of the Illinois River even in places where his samples showed few or no pollution indicators.

In Europe, the initial stress was primarily on microorganisms and results were first codified in the early years of the century by Kolkwitz and Marsson. In this "Saprobiensystem," zones of organic pollution similiar to those described by the American workers were defined and organisms were listed as characteristic of one or more zones; a recent exposition of this list is given by Kolkwitz (1950). It was then claimed that with a list of the species occurring at a particular point it was possible to allocate it to a saprobic zone. This system early met with criticism for several reasons. First, all the organisms listed occurred in natural habitats — they were not evolved in polluted water — and there was much doubt as to the placing of many of the species in the lists. The system, however, did serve to codify ecological knowledge about a long list of species along an extended trophic scale. Its weaknesses appeared to be merely due to lack of knowledge; such a rigid system took far too little account of the complexity of the reaction of organisms to their habitats. For instance, many organisms can be found, albeit rarely, in a wide range of conditions and others may occur in restricted zones for reasons that have nothing to do with water quality. We often do not know if organisms confined to clean headwaters are kept there by high oxygen content, low summer temperatures, or inability to compete with other species under other conditions. In the swift waters of Switzerland the system broke down in that some organisms appeared in more polluted zones than their position in the lists would indicate. Presumably here the controlling factor was oxygen, which was relatively plentiful in turbulent cold water. In a recent series of experiments, Zimmerman (1962) has proved that current alone has a great influence on the biota, and identically polluted water flowing at different speeds produces biotic communities characteristic of different saprobic levels. He finds this surprising, but to me it seems an expected result, for the reasons given above.

Perhaps Zimmerman's surprise reflects the deeply rooted entrenchment of the Saprobiensystem in Central Europe. Despite its obvious shortcomings it has been revised and extended. Liebmann (1951) introduced the concept of considering number as well as occurrence and very rightly pointed out that the community of organisms is what matters rather than mere species lists. But he did not stress the importance of extrinsic factors, such as current, nor that the system can only apply to organic pollution and that different types of organic pollution differ in their effects; e.g., carbohydrate solutions from paper works produce different results from those of sewage, as they contain little nitrogen and very different suspended solids. Other workers (Sladecek 1961 and references therein) have subdivided the more polluted zones, which now, instead of being merely descriptive, are considered to represent definite ranges of oxygen content, BOD, sulphide, and even *E. coli* populations. Every water chemist knows that BOD and oxygen content are not directly related and to assume that either should be more than vaguely related to the complexities of biological reactions seems to me to indicate a fundamental lack of ecological understanding. I also think it is damaging to the hope of mutual understanding between the various disciplines concerned with water quality to give the impression that one can expect to find a close and rigid relationship between water quality measurements as assessed by different sets of parameters. Inevitably these relationships vary with local conditions; what applies in a sluggish river in summer will certainly not apply to a mountain stream or even to the same river in the winter. Correlation of data, even within one discipline, needs understanding, knowledge, and judgment. Caspers and Schulz (1960) showed that the failure of the system to distinguish between waters that are naturally productive and those artifically enriched can lead to absurd results. They studied a canal in Hamburg, which because of its urban situation can only be regarded as grossly polluted. Yet it develops a rich plankton the composition of which, according to the system, shows it to be virtually clean.

Once the Saprobiensystem was accepted it was logical to attempt to reduce its findings to simple figures or graphs for presentation of results. Several such methods were developed, which are described by Tümpling (1960), who also gives the original references. In all these methods, the abundance of each species is recorded on some sort of logarithmic scale (e.g. 1 for present, 3 for frequent, 5 for common, etc.). The sums of these abundances in each saprobic level are plotted on graphs, the two most polluted zones showing as negative and others as positive. Or, the various saprobic levels are given numerical values (1 for oligosaprobic [clean], 2 for β-mesosaprobic, etc.) and the rating for each species is multiplied by its abundance number. The sum of all these products divided by the sum of all the frequencies gives a "saprobic index" for the locality. Clearly the higher this number, the worse the water quality in terms of organic pollution. In a similar way the so-called "relative Belastung" (relative load) is calculated by expressing the sums of all the abundances of organisms characteristic of the two most-polluted zones as a percentage of the sum of all abundances. Then 100 percent is completely polluted water, and clean localities will give a low number.

There are various elaborations of these methods, such as sharing of species between zones and taking account of changes in base-line as one passes downstream. None of them, however, eliminates the basic weaknesses of the system nor the fact that, as Caspers and Schulz (1960) point out, there is little agreement between the various authors in the assignment of species to the different levels. Therefore, one gains a number or a figure that looks precise and is easily understood, but it is based on very dubious foundations.

Similar systems are indigenous to North America, but were independently evolved. Wurtz (1955) and Wurtz and Dolan (1960) describe a system whereby animals are divided into sensitive-to-pollution and non-sensitive (others are ignored), and also into burrowing, sessile, and foraging species (six classes). Numbers of these species represented are plotted for each station as six histograms on the basis of percentage of total number of species. If the constitution of the fauna from control stations or from similar localities is known, it is possible to express numerically "biological depression" (i.e., percentage reduction in total number of species), "biological distortion" (i.e., change in proportions of tolerant and non-tolerant species), and "biological skewness" (changes in the ratios of the three habitat classes). Such results must, of course, be evaluated, and the definition of tolerance is quite subjective; but the method has the advantages of simplicity and dependence on control data. Like the Saprobiensystem, however, it can have no universal validity. It also suffers from the fact that it takes no account of numbers; a single specimen, which may be there by accident, carries as much weight as a dense population.

Patrick (1949) developed a similar system in which several clean stations on the water body being investigated are chosen, and the average number of species is determined occurring in each of seven groups of taxa chosen because of their supposed reaction to pollution. These are then plotted as seven columns of equal height, and data from other stations are plotted on the same scale; it is assumed that stations differing markedly from the controls will show biological imbalance in that the columns will be of very unequal heights. Number is indicated by double width in any column containing species with an unusual number of individuals. I have already questioned the usefulness of this method of presentation (Hynes 1960), and doubt whether it gives any more readily assimilable data than simple tabulation; it does however, introduce the concept of ecological imbalance.

It has long been known that ecologically severe habitats contain fewer species than normal habitats and that the few species that can survive the severe conditions are often very abundant as they lack competitors. Examples of this are the countless millions of *Artemia* and *Ephydra* in saline lakes and the *Tubifex tubifex* in foul mud. This idea has often been expressed in terms of diversity, which is some measure of numbers of species divided by number of specimens collected. Clearly, such a parameter is larger the greater the diversity, and hence the normality of the habitat. Unfortunately, though, as the number of species in any habitat is fixed, it also decreases as sample size increases so no index of diversity has any absolute value (Hairston 1959). If a definite sample size is fixed, however, in respect to numbers of organisms identified, it is possible to arrive at a constant index.

Patrick et al. (1954) in effect used this concept in a study of diatom species growing on slides suspended in water for fixed periods. They identified 8000 specimens per sample and plotted the results as number of species per interval against number of specimens per species on a logarithmic scale. This method of plotting gives a truncated normal curve for a wide variety of biotic communities. In an ordinarily diverse habitat the mode is high and the curve short; i.e., many species occur in small numbers and none is very abundant. In a severe habitat the mode is low and the curve long; i.e. there are few rare species and a few with large numbers. This, again, seems to me to be an elaborate way of presenting data and to involve a lot of unnecessary arithmetic.

Allanson (1961) has applied this method to the invertebrate faunas of streams in South Africa and has shown, as has Patrick for diatoms, that the log normal curve is flatter and longer for polluted stations; the difference, however, is not so apparent that it does not need exposition. Here, again, I would suggest that tabulated data are just as informative. Indeed I would go further and say that tabulated data are essential in the present state of our knowledge. We are learning as we go along and if the details of the basic findings are concealed by some sort of arithmetical manipulation they cannot be re-interpreted in the light of later knowledge, nor are they preserved in the store of human knowledge. This point becomes particularly clear when one examines some of the early studies that include tables. Butcher (1946) requotes a considerable amount of data he collected from studies of various English rivers during the thirties; they are not only clear and easy to follow, but they are also informative about the generalities of pollution in a way that data quoted only within the confines of some particular system are not.

Simple tabulation of biological data in relation to water quality, either in terms of number of organisms, percentage composition of the biota, some arbitrary abundance scale, or as histograms, has been effectively practiced in many parts of the world: in America (Gaufin and Tarzwell 1952, Gaufin 1958), Africa (Harrison 1958 and 1960, Hynes and Williams 1962), Europe (Albrecht 1954, Kaiser 1951, Hynes 1961, Hynes and Roberts 1962), and New Zealand (Hirsch 1958) to cite a few. These tabulated data are easy to follow, are informative to the expert reader, and conceal no facts. Although the non-biologist may find them tedious, he need only read the explanatory paragraphs. It is a delusion to think that it is possible to reduce biological data to simple numerical levels. At best, these can only be produced for limited situations and even then they need verbal exposition; at worst, they give a spurious impression of having absolute validity.

My final point in this section concerns comparisons. It is claimed that the German system, in effect, measures an absolute state, a definite level of water quality. We have seen that this is not a tenable claim. In the other systems, by and large, the need to establish local control stations at which to measure the normal or "natural" biotic conditions is accepted, and then other areas are compared with this supposed norm. This is, of course not always possible as there may remain no unaffected area, or no unaffected area that is, with respect to such factors as current, nature of substratum etc., sufficiently similar to act as a base-line for data. Nevertheless, basically, these systems can be used to compare stations and thus to assess *changes* in water quality. In doing this, they can all be used more or less successfully, but I maintain that a table is just as useful as an elaborate analysis, and I believe that the table should be included with whatever is done. For a particular situation, however, it is often possible to distill the data into a single figure as a measure of similarity between stations.

Burlington (1962 and Dean and Burlington 1963) has recently proposed an entirely objective means of doing this, which involves simple arithmetical manipulation. In his system a "prominence value" is calculated for each species at each station. This is a product of its density and some function of its frequency in samples, but the details of this calculation can be altered to suit any particular situation. Then a coefficient of similarity between each pair of stations can be calculated by dividing twice the sum of the lower prominence values of taxa that the two stations have in common by the sum of all the prominence values of both stations. Identical stations will then have a coefficient of similiarity of 1.00; this coefficient will be lower the more different the stations are from one another. This is an easy way to compare stations in an entirely unbiased way and as such may satisfy the need for numerical exposition; however, it tells one nothing about why the localities are different and like all the other more or less numerical methods of presenting data has no absolute value. Moreover, it still leaves unanswered the fundamental question of how different is "different?"

THE PROBLEMS OF SAMPLING

The systems outlined above are all based on the assumption that it is possible to sample an aquatic habitat with some degree of accuracy; this is a dubious assumption, however, when applied to biological data. From what has been said about the complexity of biological reactions to the various factors in the environment, and from the obvious fact that rivers especially are a mosaic of microhabitats, it is clear that to achieve numerical accuracy or even some limits of confidence considerable numbers of samples need to be taken. Indeed, even in so apparently unvaried a habitat as a single riffle, Needham and Usinger (1956) showed that a very large number of samples would be necessary to give significant numerical data.

There is a limit to the number of samples that can reasonably be taken and, anyway, it is desirable to sample many different types of habitat so as to get as broad as possible an estimate of the biota. This is the more recent approach of most of the workers in Central Europe, who have been content to cite abundances on a simple relative but arbitrary scale and to convert this to figures on some sort of logarithmic scale for use in calculations. An alternative is to express the catch in terms of percentage composition, but this has the disadvantage that micro- and macro-organisms cannot be expressed on the same scale as they are obtained by different collecting techniques. Also, of course, implicit in this approach is the assumption that the sampling is reasonably representative. Here again we run into the need for knowledge and expertise. In collection as well as in interpretation, the expert is essential. Biological sampling, unlike the simple, or fairly simple, filling of bottles for chemical analysis or the monitoring of measuring equipment, is a highly skilled job and not one to be handed over to a couple of vacationing undergraduates who are sent out with a Surber sampler and told to get on with it. This point has also been made by other biologists, e.g., Patrick (1961) who stresses the need for skilled and thorough collecting even for the determination of a species list.

Alternatively we can use the less expert man when concentrating on only part of the habitat, using, say, microscopical slides suspended in the water to study algal growth. This method was extensively used by Butcher (1946), and Patrick et al. (1954) who studied diatoms in this way. This gives only a partial biological picture, but is useful as a means of monitoring a stretch of water where it is possible that changes might occur. It is a useful short-hand method, and as such is perhaps comparable to studying the oxygen absorbed from potassium permanganate instead of carrying out all the usual chemical analyses on water. A short method of this kind may serve very well most of the time, but, for instance, would not be likely to detect an insecticide in concentrations that could entirely eliminate arthropods and hence fishes by starvation.

It is possible to work out biological monitoring systems for any specific purpose. The simplest of these is the cage of fish, which, like a single type of chemical analysis, can be expected to monitor only one thing — the ability of fish to live in the water — with no information on whether they can breed or whether there is anything for them to eat. Beak et al. (1959) describe a neat way in which the common constituents of the bottom fauna of Lake Ontario can be used to monitor the effluents from an industrial site. Obviously there is much room for such ingenuity in devising biological systems for particular conditions, but this is perhaps outside the scope of this meeting.

CONCLUSIONS

It may appear from the previous sections that my attitude to this problem is entirely obstructionist. This is far from being so. Water quality is as much biological phenomenon as it is a chemical or physical one; often what we want to know about water is almost exclusively biological — will it smell nasty, is it fit to drink, can one bathe in it, etc.? I suggest, therefore, that it is desirable to organize water monitoring programs that will tell one what one wants to know. There is no point in measuring everything biological, just as there is no point in performing every possible chemical analysis; what is measured should be related to local conditions. It would be a waste of time to measure oxygen content in a clean mountain stream; we know it to be high, and it becomes worth measuring only if we suspect that it may have been lowered by pollution. Similarly, there is little point in studying the plankton in such a stream; we know it only reflects the benthic flora. In a lake or in a slow river, on the other hand, if our interest in the water lies in its potability, records of the plankton are of considerable importance as changes in plankton are, in fact, changes in the usability of the water.

For long-term studies, especially for the recording of trends or changes induced by pollution, altered drainage, agricultural poisons, and other havoc wrought by man, one can expect informative results from two principal techniques: First, we can study microscopic plant and animal growth with glass slides placed in the water for fixed periods; second, we can obtain random samples of the benthic fauna. The algae and associated microfauna tell one a good deal about the nutrient condition of the water and the changes that occur in it, and the larger benthic fauna reveal changes in the trophic status, siltation due to soil erosion, effects of insecticides and other poisons, etc.

The study of growths on glass slides is reasonably skilled work, but can easily be taught to technicians; like chemical monitoring, such study needs to be done fairly often. Sampling the benthos is more difficult and, as explained above, needs expert handling; unlike most other monitoring programs, however, it need be done only infrequently, say, once or twice a year. Inevitably sampling methods will vary with type of habitat; in each case, the question will arise as to whether it is worth looking at the fish also. It is here that the biologist must exercise judgment in devising and carrying out the sampling program.

Judgment is also needed in the interpretation of the data. It is for this reason I maintain that it should all be tabulated so that it remains available for reassessment or comparison with later surveys. If need be, some sort of numerical format can be prepared from the data for *ad hoc* uses, but it should never become a substitute for tabulations. Only in this way can we go on building up our knowledge. Perhaps some day we shall be able to pass all this information into a computer, which will then be able to exercise better judgment than the biologist. I hope this will happen, as computers are better able to remember and to cope with complexity than men. It will not, however, pension off the biologist. He will still be needed to collect and identify the samples. I cannot imagine any computer wading about on rocky riffles nor persuading outboard motors and mechanical graps to operate from the unstable confines of small boats. We shall still need flesh and blood biologists long after the advent of the hardware water chemist, even though, with reference to my earlier analogy, a Tokyo University computer recently outpointed 10 veteran medicals in diagnosing brain tumors and heart disease. It should be pointed out, however, that the computer still had to be fed with information, so we are still a long way from the hardware general practitioner. I believe though that he is likely to evolve before the hardware biologist; after all, he studies only one animal.

REFERENCES

Albrecht, M.-L. (1954). Die Wirkung der Kaliabwässer auf die Fauna der Werra und Wipper. *Z. Fisch. N. F. 3*:401-26.

Allanson, B. R. (1961). Investigations into the ecology of polluted inland waters in the Transvaal. Part I. *Hydrobiologia 18*:1-94.

Beak, T. W., de Courval, C. and Cooke, N. E. (1959). Pollution monitoring and prevention by use of bivariate control charts. *Sew. Industr. Wastes 31*:1383-94.

Beck, Wm. M., Jr. (1957). The use and abuse of indicator organisms. *Transactions of a Seminar on Biological Problems in Water Pollution*. Cincinnati.

Burlington, R. F. (1962). Quantitative biological assessment of pollution. *J. Wat. Poll. Contr. Fed. 34*:179-83.

Butcher, R. W. (1946). The biological detection of pollution. *J. Inst. Sew. Purif. 2*:92-7.

Caspers, H. and Schulz, H. (1960) Studien zur Wertung der Saprobiensysteme. *Int. Rev. ges. Hydrobiol. 45*:535-65.

Dean, J. M. and Burlington, R. F. (1963). A quantitative evaluation of pollution effects on stream communities. *Hydrobiologia 21*:195-9.

Gaufin, A. R. (1958). The effects of pollution on a midwestern stream. *Ohio J. Sci. 58*:197-208.

Gaufin, A. R. and Tarzwell, C. M. (1952). Aquatic invertebrates as indicators of stream pollution. *Pub. Hlth. Rep. 67*:57-64.

Hairston, N. G. (1959). Species abundance and community organization. *Ecology 40*:404-15.

Harrison, A. D. (1958). The effects of sulphuric acid pollution on the biology of streams in the Transvaal, South Africa. *Verh. int. Ver. Limnol. 13*:603-10.

Harrison, A. D. (1960). The role of river fauna in the assessment of pollution. *Cons. sci. Afr. Sud Sahara Pub. 64*:199-212.

Hirsch, A. (1958). Biological evaluation of organic pollution of New Zealand streams. *N. Z. J. Sci. 1*:500-53.

Hynes, H. B. N. (1960). *The biology of polluted waters.* Liverpool.

Hynes, H. B. N. (1961). The effect of sheep-dip containing the insecticide BHC on the fauna of a small stream. *Ann. trop. Med. Parasit. 55*:192-6.

Hynes, H. B. N. and Roberts, F. W. (1962). The biological effects of detergents in the River Lee, Hertfordshire. *Ann. appl. Biol. 50*:779-90.

Hynes, H. B. N. and Williams, T. R. (1962). The effect of DDT on the fauna of a Central African stream. *Ann. trop. Med. Parasit. 56*:78-91.

Illies, J. (1961a). *Die Lebensgemeinschaft des Bergbaches.* Wittenberg-Lutherstadt.

Illies, J. (1961b). Versuch einer allgemeiner biozönotischen Gliederung der Fliessgewässer. *Int. Rev. ges. Hydrobiol. 46*:205-13.

Kaiser, E. W. (1951). Biologiske, biokemiske, bacteriologiske samt hydrometriske undersogelser af Poleaen 1946 og 1947. *Dansk. Ingenforen. Skr. 3*:15-33.

Kolkwitz, R. (1950). Oekologie der Saprobien. Uber die Beziehungen der Wasserorganismen zur Ummelt. *Schr. Reihe ver Wasserhyg. 4*:64 pp.

Liebmann, H. (1951). *Handbuch der Frischwasser und Abwasserbiologie.* Munich.

Needham, P. R. and Usinger, R. L. (1956). Variability in the macrofauna of a single riffle in Prosser Creek, California, as indicated by the Surber sampler. *Hilgardia 24*:383-409.

Patrick, R. (1949). A proposed biological measure of stream conditions, based on a survey of the Conestoga Basin, Lancaster County, Pennsylvania. *Proc. Acad. Nat. Sci. Phila. 101*:277-341.

Patrick, R. (1961). A study of the numbers and kinds of species found in rivers in Eastern United States. *Proc. Acad. Nat. Sci. Phila. 113*:215-58.

Patrick, R., Hohn, M. H. and Wallace, J. H. (1954). A new method for determining the pattern of the diatom flora. *Not. Nat. Phila. Acad. Sci. 259*:12 pp.

Richardson, R. E. (1929). The bottom fauna of the middle Illinois River, 1913-1925; its distribution, abundance, valuation and index value in the study of stream pollution. *Bull. Ill. nat. Hist. Surv. 17*:387-475.

Sladecek, V. (1961). Zur biologischen Gliederung der höheren Saprobitätsstufen. *Arch. Hydrobiol.* *58*:103-21.

Tümpling, W. v. (1960). Probleme, Methoden und Ergenbnisse biologischer Gütenuntersuchungen an Vorflutern, dargestellt am Beispiel der Werra. *Int. Rev. ges. Hydrobiol.* *45*:513-34.

Whipple, G. C., Fair, G. M. and Whipple, M. C. (1947). *The microscopy of drinking water*. New York.

Wurtz, C. B. (1955). Stream biota and stream pollution. *Sew. industr. Wastes* 27:1270-8.

Wurtz, C. B. and Dolan, T. (1960). A biological method used in the evaluation of effects of thermal discharge in the Schuylkill River. *Proc. Ind. Waste Conf. Purdue*, 461-72.

Zimmerman, P. (1962). Der Einfluss auf die Zusammensetzung der Lebensgemeinschaften in Experiment. *Schweiz. Z. Hydrol.* *24*:408-11.

Dr. Werner Stumm
Associate Professor of Applied Chemistry
Harvard University
Cambridge, Massachusetts

SUMMARY

This paper considers some of the chemical reactions that may, at least partially, determine the composition of fresh water. Examples are given that demonstrate how elementary principles of physical chemistry can aid in the identification of various interrelated variables that determine the mineral relations in natural water systems. In a hypothetical experiment, a unit volume of "natural" fresh water was prepared by sequentially mixing with distilled water some of the pertinent constituents, starting with more abundant ones. After each addition, the equilibrium composition was calculated and compared with the composition of that in a real natural water system. Throughout the experiment, standard reference tables on the energies or relative stabilities of the various compounds were used. The stability relationships are shown in simple graphs.

CHEMISTRY OF NATURAL WATERS IN RELATION TO WATER QUALITY

Natural waters acquire their chemical characteristics through direct solution and chemical reactions with solids, liquids, and gases with which they have come in contact during the various parts of the hydrological cycle. The final composition of a natural water is the result of a great variety of chemical, physical, and biological reactions.

This paper considers some of the chemical reactions that may, at least partially, determine the composition of fresh water. Obviously, this discussion of the physical chemistry of natural waters cannot be comprehensive. The author has concentrated on some examples that are in his opinion best suited methodologically and didactically to demonstrate how elementary principles of physical chemistry can aid in identifying the various interrelated variables that determine the mineral relations in natural water systems. In writing this discussion, the author could not avoid being strongly influenced by Sillen's excellent paper on the "Physical Chemistry of Sea Water."[1]

THE MODEL

Since it is not possible to evaluate all the various chemical process combinations and the various environmental factors, e.g., mineralogical and geological environment, rate of water circulation, biological activity, temperature and pressure, etc., a simplified model will be chosen. In a hypothetical experiment, we shall prepare a unit volume of "natural" fresh water by sequentially mixing with distilled water some of the pertinent constituents, starting with the more abundant ones. After each addition the equilibrium composition will be calculated. For this calculation we will use free energy data (equilibrium constants, redox potentials) found in standard references.[2,3] The composition of the water in a model at equilibrium will be compared with the composition of that in a real natural water system.

LIMITATIONS OF THE MODEL

This hypothetical experiment is didactical. The sequence of addition of chemicals to the pure water is not an attempt to follow the geological history and is thus rather arbitrary. The comparison between the equilibrium model and the real system must take into consideration that a true equilibrium is not necessarily attained in all respects in the real system. In a natural body of water only the upper layers are in contact with the atmosphere and only the deepest layers are in contact with the uppermost layers of the sediments. The mixing in the real system is further impaired by density stratification due to vertical temperature differences. On the other hand, the real systems are subject to periodic overturns; geological time spans have elapsed, and therefore, reactions that reach equilibrium very slowly in the laboratory may have come nearer equilibrium in real systems. We must also be aware that biologically mediated reactions, e.g., photosynthesis and respiration, can lead to significant localized disturbances of the equilibrium composition.

The results obtained for the equilibrium model, of course, contain only that information (free energy data for the species considered) that has been used for their computation. The available free energy values or equilibrium constants are frequently not known with sufficient precision, some data are lacking, and occasionally we may overlook a pertinent species. In view of these inadequacies, not much attention has been paid to activity corrections and all calculations are based on 25°C. Consequently, in most instances the results obtained represent an oversimplified picture. Nevertheless, it is gratifying to see that the predictions are frequently in reasonable accord with observed behavior in real systems.

The comparison between natural systems and their idealized counterparts is an essential prerequisite to isolation of the variables responsible for observed mineral relations. The equilibrium calculations and the comparison of the results with those of the real systems will permit us to make some speculation on the type of solid phases and dissolved species one may expect in fresh water systems. The value of the model thus lies primarily in providing an aid for the interpretation of observed facts. Discrepancies between equilibrium predictions and chemical data of the real systems can give us valuable insight into those circumstances where chemical reactions are not sufficiently understood, where non-equilibrium conditions prevail in the real systems, or where the analytical data available are not sufficiently accurate or specific.

MAJOR COMPONENTS

SILICON AND ALUMINUM

At first sight, it might appear somewhat puzzling that we start our imaginary experiment with these two elements as major constituents. It has frequently been assumed that both silicon and aluminum do not hold an important position in mineral water quality relations. This is only true if we consider waters in isolation from their natural environment. But it is so frequently forgotten by those who deal with water resources that every lake and every body of natural water has a bottom (igneous and sedimentary rocks). Dissolved mineral matter originates in the crustal materials of the earth; water disintegrates mineral rocks by erosion and weathering and acts as a solvent on almost all of them. Goldschmidt[4] has estimated that for each kilogram of ocean water some 600 grams of primary igneous rock must have been decomposed. Similar estimates cannot be made in such a general way for fresh waters, but it might be safe to assume that practically every ground and surface water has been in extensive and intimate contact with sedimentary rocks. A qualitative illustration of such rock mineral and water interaction is given by records of the U. S. Geological Survey,[5] which show that from 70 to 86 tons of soluble matter were carried, on the average in 1950, from each square mile of drainage area of the James River above Richmond, Virginia; the Iowa River above

Iowa City; and the Colorado River above Grand Canyon, Arizona. Higher rates were observed for streams draining limestone terranes. Silicon and aluminum are, besides oxygen, the most abundant elements in igneous and sedimentary rocks. Although relatively small amounts of these elements become homogeneously dissolved in water, the various equilibria for heterogeneous chemical reactions between the solid and solution phases are probably of utmost significance in the chemistry of natural waters. Unfortunately, much of the mineral and solution chemistry of these elements is not yet well understood.

In making our artificial body of water, we add solid SiO_2 to pure water. Since not very much of this SiO_2 will become dissolved, it is not critical how much SiO_2 we add as long as we maintain an excess of solid SiO_2. It might appear reasonable to add about 2 mole of SiO_2 per liter of pure water. The various reactions that can occur are listed with their respective equilibrium constants in Table 1.* Reactions 1 and 2 describe the solubility equilibrium of SiO_2. It is obvious that quartz is the thermodynamically stable form of SiO_2, whereas amorphous SiO_2 is metastable and about 10 times more soluble than quartz. (Ortho-)silicic acid, $Si(OH)_4$, is a very weak acid (reactions 3 and 4); its conjugate monoprotic and diprotic bases, the silicates $SiO(OH)_3^-$ and $SiO_2(OH)_2^{-2}$, are not important constituents in the common pH range of natural waters (pH 6 to 9). Thus, for the dissolution of SiO_2, reactions 1 and 2 primarily have to be considered. In our mixture, we will find about 2×10^{-4}M dissolved $Si(OH)_4$ if we use quartz (or sandstones), and about 10 times more if we use amorphous silica as the source of SiO_2. Natural waters can thus contain up to approximately 56 milligrams of dissolved silica per liter, if we assume that the amorphous forms of SiO_2 are the major source of silica in natural waters. In most natural waters concentrations range from 0.5 to about 15 milligrams per liter, although concentrations up to 50 milligrams per liter are not uncommon. The solubility of SiO_2 as $Si(OH)_4$ increases with increasing temperature; thus, hot springs frequently have higher dissolved silica than cold waters.

Table 1 — SiO_2 - Equilibria (Reference 6)

Reaction No.	Reaction	log K
1	SiO_2 (quartz) $+ H_2O = Si(OH)_4$	— 3.7
2	SiO_2 (amorph) $+ H_2O = Si(OH)_4$	— 2.7
3	$Si(OH)_4 = [SiO(OH)_3]^- + H^+$	— 9.5
4	$[SiO(OH)_3]^- = [SiO_2(OH)_2]^{-2} + H^+$	—12.6
5	$4\,Si(OH)_4 = [Si_4O_6(OH)_6]^{-2} + 4\,H_2O + H^+$	—12.6

On the basis of the data given in Table 1, it must be concluded, contrary to earlier beliefs, that silica in water does not occur as a colloid. Most natural waters (below pH 9) do not contain silicate anions in appreciable concentrations. Therefore, dissolved silica, under these conditions, is not a part of the operationally determined alkalinity of a water, nor does the dissolved $Si(OH)_4$ have any marked influence on the buffer capacity of fresh waters.[7]

---

* For most of the reactions listed in the tables in this paper, different authors may have determined different equilibrium constants. The constants given in these Tables have generally been selected from tabulated values given in references 2 and 3. Only in those instances where other sources have been used will special reference be given. All constants given apply to 25°C and do not always refer to the proper ionic strength of natural waters (5×10^{-4} to 5×10^{-3}).

Figure 1 gives a solubility diagram for SiO_2 (total soluble silica as a function of pH). The dissolution of SiO_2 becomes significant at very high pH values (water glass) (reactions 3, 4, and 5). The polymerization of $Si(OH)_4$ to tetrameric silicate (reaction 5) occurs only under alkaline conditions ($pH > 10$). If alkaline concentrated solutions that contain polymeric silicates are acidified to lower pH values, the solubility of silica is exceeded and SiO_2 precipitates. Within neutral and slightly alkaline pH ranges, relatively stable negatively charged collodial dispersions of SiO_2 (activated silica) are formed.

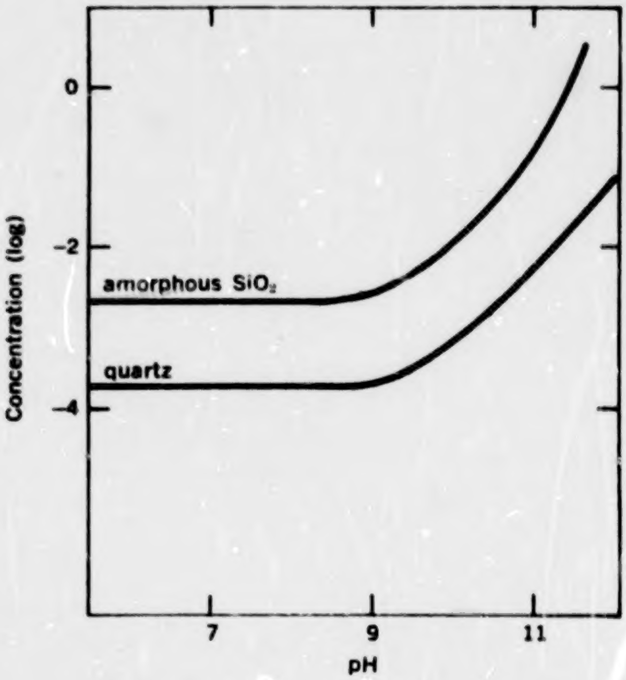


Figure 1 — Solubility of Quartz and Colloidal SiO_2.

Besides SiO_2, various silicate minerals, metal silicates, and clays are important components of mineral rocks that represent sources of dissolved material in water.

We now add about 1 mole of $Al(OH)_3$ per 1 liter to our water. As a first approach, we might ask ourselves how much of the $Al(OH)_3$ would become soluble. The chemistry of Al(III) has not been elucidated in great detail, but some of the more recent equilibrium information is summarized in Table 2.

Table 2 — Hydrolysis and Solubility Equilibria of Aluminum

Reaction No.	Reaction	log K
6	$Al^{+3} + H_2O = [AlOH]^{+2} + H^+$	— 5.0
7a	$Al^{+3} + 3\,H_2O = Al(OH)_3(s) + 3\,H^+$	— 9.1
7b	$Al(OH)_3(s) = Al^{+3} + 3\,OH^-$	—32.9
8	$Al(OH)_3(s) + H_2O = [Al(OH)_4]^- + H^+$	—12.7
9	$6\,Al^{+3} + 15\,H_2O = [Al_6(OH)_{15}]^{+3} + 15\,H^+$	—47
10	$8\,Al^{+3} + 20\,H_2O = [Al_8(OH)_{20}]^{+4} + 20\,H^+$	——

On the basis of this information, a solubility diagram has been constructed (Figure 2). From this, it is evident that Al^{+3} is very easily hydrolyzed to various hydroxide complexes. Aquo-aluminum ion is an acid that exhibits acidity similar to that of acetic acid. There is some uncertainty about the various hydrolysis products that might occur in the slightly acid to neutral pH range. Although the behavior of dilute Al(III) solutions can be reasonably well interpreted on the basis of reaction 6, the potentiometric investigations of Brosset and co-workers[8] on the reaction of Al^{+3} ion with water in the presence of various concentrations of OH– ions have revealed that the monomeric hydrolysis species $Al(OH)^{+2}$ is not the main hydrolysis product if it exists at all. Brosset was able to interpret his data by postulating a soluble polymeric aluminum hydroxo complex with a stoichiometric ratio of OH– to Al(III) of 2.5. He suggested $[Al_6(OH)_{15}]^{+3}$ as the most likely structure. On the basis of colloid chemical investigations, Matijevic and co-workers[9] postulate $[Al_8(OH)_{20}]^{+4}$ to be the most likely formula. From the data given in Table 2 and Figure 2, it is evident that $Al(OH)_3(s)$ exhibits amphoteric properties. The solubility of $Al(OH)_3$ increases in the acid and alkaline range. With increasing pH, more hydroxide ions are coordinated to the aluminum, and soluble aluminate, $[Al(OH)_4]^-$ or $[Al_2(OH)_8]^{-2}$, is formed.

Figure 2 indicates that in the common pH range of natural waters the predominant soluble aluminum species appears to be aluminate, $(Al(OH)_4]^-$. Total soluble Al(III) in equilibrium with $Al(OH)_8(s)$ amounts to approximately 10^{-6}M at pH 7 and approximately 10^{-5}M at pH 8. At these pH values, Al^{+3} would account for only about 10^{-12}M and 10^{-15}M, respectively. Total soluble Al(III) in most natural waters should vary between about 30 micrograms per liter (pH 7) and 300 micrograms per liter (pH 8). Little reliable analytical data on the Al(III) content of natural waters are available for comparison with these calculated equilibrium results.

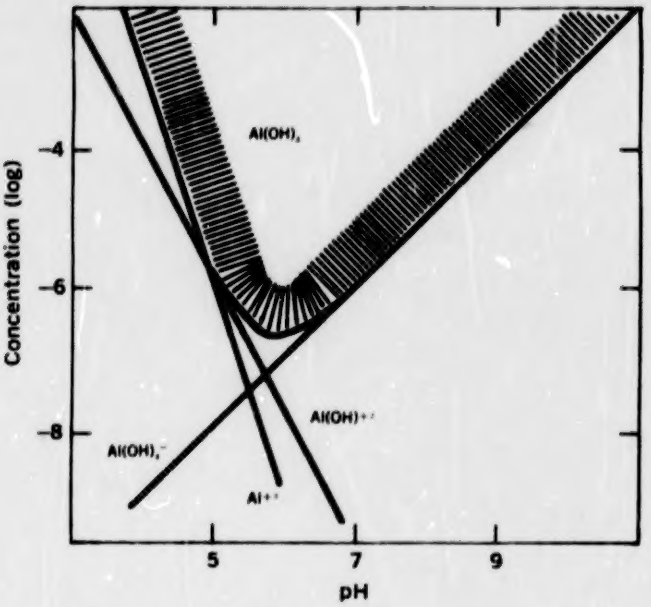


Figure 2 — Solubility of Aluminum Hydroxide.

All the $Al(OH)_8$ that has been added to our mixture will, however, ultimately react with SiO_2 to form aluminum silicate minerals such as kaolinite, $Al_2Si_2O_5(OH)_4(s)$. Since SiO_2 is in excess of $Al(OH)_3$, the silicic acid content of the solution will not change.

Aluminum silicates like kaolinite can rearrange their structure in such a way that Mg^{+2} (or Ca^{+2}) may substitute for aluminum in its octahedral coordination arrangement. In a similar way aluminum may replace silica in its tetrahedral structure. Negatively charged aluminum silicate frameworks with layer structure (clays) are built up in such a manner. Because of their electronegative nature, these clays exhibit cation-exchange phenomena. Although it has not been established which solid clay phases represent true equilibrium states, we must take into account that a great variety of clays are encountered as metastable solid phases in aquifers, in sediments of surface waters, and in suspension. Ion-exchange equilibria between dissolved constituents of natural waters and clays and minerals with which these waters come into contact influence the concentration of H^+ and other cations. Sillen[1] has identified hetereogeneous silicate equilibria as comprising the principal pH buffer system in oceanic waters. A plausible reaction scheme for strongly pH dependent silicate equilibria has been given by Sillen[1]:

$$3Al_2Si_2O_5(OH)_4(s) + 4SiO_2(s) + 2K^+ + 2Ca^{+2} + 9H_2O = \\ 2KCaAl_3Si_5O_{16}(H_2O)_6(s) + 6H^+ \qquad (11)$$

Although equilibrium relationships of such reactions are not yet well understood, it is obvious that exchange reactions such as given in equation 11 must exert considerable influence upon the hydrogen ion concentration of natural waters.

CALCIUM CARBONATE

We will now add to our mixture $CaCO_3$ in the proportion of about 0.5 mole per liter, thus introducing Ca^{+2} and carbon simultaneously.

Since the previous additions of $Al(OH)_3$ and SiO_2 did not have any appreciable influence upon the pH of the solution (there has been a very slight reduction in pH in this unbuffered system) and since the already dissolved species will have no influence upon the $CaCO_3$ solubility equilibrium, our problem of equilibrium calculation can essentially be reduced to that of pure water being in contact and in equilibrium with solid $CaCO_3$. The equilibria that have to be considered are listed in Table 3.

Table 3 — $CaCO_3$ and Carbonate Equilibria

Reaction No.	Reaction	log K
12	$CO_2(g) + H_2O = H_2CO_3^*$	-1.5 (K)
13	$H_2CO_3^* = HCO_3^- + H^+$	-6.3 (K_1)
14	$HCO_3^- = CO_3^{-2} + H^+$	-10.3 (K_2)
15a	$CaCO_3(s) = Ca^{+2} + CO_3^{-2}$	-8.3 (K_s)
15b	$CaCO_3(s) + H^+ = Ca^{+2} + HCO_3^-$	$+2.0$ (K_s/K_2)
15c	$CaCO_3(s) + H_2CO_3^* = Ca^{+2} + 2\,HCO_3^-$	-4.3 (K_sK_1/K_2)
16	$CaCO_3(aq) = Ca^{+2} + CO_3^{-2}$	-3.0 (?)

Remarks: $H_2CO_3^*$ refers to the sum of dissolved CO_2 and H_2CO_3. In order to simplify the writing of the equations in the text the following terms are introduced to define the total concentration of dissolved carbonic species, C_T; alkalinity, [Alk].

$$C_T = [H_2CO_3^*] + [HCO_3^-] + [CO_3^{-2}] \qquad (17)$$

$$[\mathrm{Alk}] \quad = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^-}] + [\mathrm{OH^-}] - [\mathrm{H^+}] \tag{18}$$

$$[\mathrm{Acidity}] = 2[\mathrm{H_2CO_3^*}] + [\mathrm{HCO_3^-}] + [\mathrm{H^+}] - [\mathrm{OH^-}] \tag{19}$$

The following abbreviations are derived from Equations 13, 14, and 17:

$$\alpha_0 = [\mathrm{H_2CO_3^*}]/C_T = 1/(1 + K_1/[\mathrm{H^+}] + K_1K_2/\,[\mathrm{H^+}]^2) \tag{20}$$

$$\alpha_1 = [\mathrm{HCO_3}-]/C_T = 1/(1 + [\mathrm{H^+}]/K_1 + K/_2\,[\mathrm{H^+}]) \tag{21}$$

$$\alpha_2 = [\mathrm{CO_3^{-2}}]/C_T = 1/(1 + [\mathrm{H^+}]/K_2 + [\mathrm{H^+}]^2/K_1K_2) \tag{22}$$

Furthermore, the ion product of water $[\mathrm{H^+}]\,[\mathrm{OH^-}] = K_w$ is taken as 10^{-14}.

CASE 1: SYSTEM CLOSED TO THE ATMOSPHERE

As a first approximation, we assume that our system is not exposed to the atmosphere and we treat $H_2CO_3^*$ as a non-volatile acid. Under these circustances all Ca^{+2} that becomes dissolved must equal in concentration the sum of the dissolved carbonic species:

$$[\mathrm{Ca^{+2}}] = C_T \tag{23}$$

Furthermore, the solution must fulfill the condition of electroneutrality:

$$2[\mathrm{Ca^{+2}}] + [\mathrm{H^+}] = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^{-2}}] + [\mathrm{OH^-}] \tag{24a}$$

or

$$2[\mathrm{Ca^{+2}}] = [\mathrm{Alk}] \tag{24b}$$

i.e., in such a solution the calcium hardness is equal (equivalent) to the alkalinity.

Since, according to equations 15, 22, and 23, $[Ca^{+2}]$ or C_T is equal to $(K_s/\alpha_2)^{0.5}$, we can, considering equation 21, rewrite the electroneutrality condition 24 in the following way*:

$$(K_s/\alpha_2)^{0.5}[\,2 - \alpha_1 - 2\alpha_2] + [\mathrm{H^+}] - K_w/\,[\mathrm{H^+}] = 0 \tag{25}$$

This equation can most conveniently be solved for $[H^+]$ by trial and error. For the given $[H^+]$, the equilibrium concentration of the additional dissolved species can readily be accomplished. The result of such computation gives:

$$\mathrm{pH} = 9.9;\ [\mathrm{Ca^{+2}}] = 1.2 \times 10^{-4};\ [\mathrm{HCO_3^-}] = 9 \times 10^{-5};$$
$$[\mathrm{CO_3^{-2}}] = 4 \times 10^{-5};\ [\mathrm{H_2CO_3^*}] = 2.5 \times 10^{-8};$$
$$[\mathrm{Alk}] = 2.4 \times 10^{-4};\ [\mathrm{Acidity}] = 0.$$

INFLUENCE OF ACID AND BASE

The equilibrium mixture ($CaCO_3$ + water) thus obtained is not well buffered (we disregard for the moment the influence of SiO_2 and $Al(OH)_3$), and small quantities of acids or bases will change the pH and the solubility relations. We might visualize that such pH changes could occur, for example, upon addition of acid or base waste constituents, through dissolution of volcanic HCl or through the influence of biological reactions (e.g., H^+ addition as a result of nitrification or OH^- addition as a result of denitrification). From a computational point of view the problem is analogous to the titration of a $CaCO_3$ suspension with strong acid, C_A, or strong base, C_B. Such acid or base addition will shift the electroneutrality condition expressed in equation 24 to the following:

---

* In this and many of the subsequent equations, some of the terms are (even in very exact calculations) negligible. Generally, mathematically exact equations are given. This "precision" might appear to be in contrast with the many otherwise uncertain factors involved in these calculations, but the author believed it necessary to emphasize that the quantitative evaluation of the systematic relations that determine equilibrium concentrations of a solution constitutes a purely mathematical problem that is, without the need for introducing a priori assumptions, subject to exact and systematic treatment.[10] A relatively simple way to survey the interrelationships of the equilibrium concentrations of the individual solute species consists of a simultaneous graphical representation of all the requisite equations (see Figures).[11, 12]

$$C_A - C_B = (K_s/\alpha_2)^{0.5}(2 - \alpha_1 - 2\alpha_2) + [H+] - K_w/[H+] \qquad (26)$$

With the help of this equation, it is always possible to compute the quantity of C_A or C_B needed per liter of water in contact and in equilibrium with solid $CaCO_3$ to reach a given pH value (Figure 3). Of course, the addition of acid and the resultant lowering of pH will lead to an increase in dissolved Ca^{+2} and carbonic constituents (Ca hardness > [Alk]), whereas base addition will result in deposition of $CaCO_3$ (Ca hardness < [Alk]). Under our assumptions the condition of equation 23 still holds; thus, the pH dependence of soluble Ca^{+2} and of the sum of carbonic species, C_T, is determined by

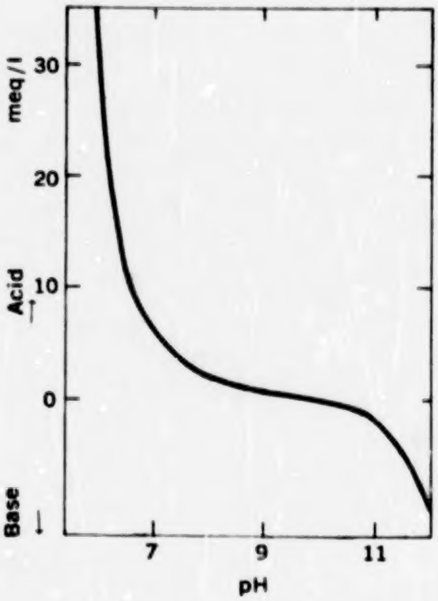


Figure 3 — Titration of $CaCO_3$ Suspension With Acid and Base.

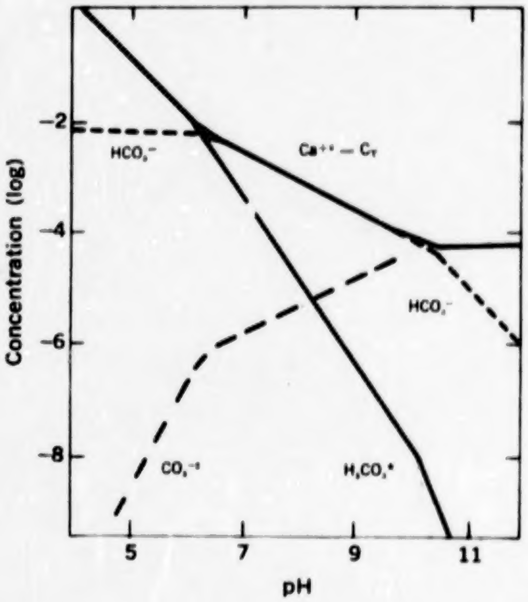


Figure 4 — Dissolved Species of a $CaCO_3$ Suspension.

$$[Ca^{+2}] = C_T = (K_s/\alpha_2)^{0.5} \qquad (27)$$

Equilibrium concentrations for Ca^{+2}, HCO_3^-, CO_3^{-2}, H_2CO_3, and alkalinity are depicted in Figure 4. (Equation 26 and Figures 3 and 4 represent the essential quantitative principles involved in the Ca^{+2} removal by lime-soda softening.)

CASE 2: SYSTEM OPEN TO ATMOSPHERE

In our calculations so far, we have neglected the influence of atmospheric CO_2 and have treated $H_2CO_3^*$ as a non-volatile acid. In order to approach more realistic conditions, we open our system to atmospheric CO_2 and we assume that the partial pressure of CO_2 is equal to approximately 3 x 10^{-4} atmosphere. On the basis of Henry's law (equation 12), the equilibrium concentration of $H_2CO_3^*$ is given by approximately $Kp^{CO_2} = 10^{-5}$. The electroneutrality condition of equation 24a still applies; $[Ca^{+2}]$ is no longer equal to C_T, but equation 24b is still valid, i.e., the calcium hardness is equivalent to alkalinity. The equilibrium condition of equation 24 can be rewritten in the following form:

$$2\,\frac{K_s\alpha_0}{\alpha_2 Kp_{co_2}} = \frac{Kp_{co_2}}{\alpha_0} \times (\alpha_1 + 2\alpha_2) + [OH^-] - [H+] \qquad (28)$$

Solution of this equation gives:

$[Ca^{+2}] = 5 \times 10^{-4}$; $[CO_3^{-2}] = 1 \times 10^{-5}$;
$[HCO_3^-] = 10^{-3}$; pH $= 8.4$; $[H_2CO_3^*] = 10^{-5}$;
$[Alk] = 10^{-3}$

Comparing this result with that of Case 1, we see that the influence of atmospheric CO_2 has depressed the pH markedly and that the concentration of $[Ca^{+2}]$ and [Alk] has been raised to values very representative of those in natural waters.

CASE 3: WATER ISOLATED FROM SOLID $CaCO_3$

In a water isolated from its sediments and mineral rocks, such as water in epilimnetic layers of a lake, or in samples brought to the laboratory, the presence of $H_2CO_3^*$, HCO_3^-, and CO^{3-2} is primarily responsible for the maintenance of near neutral pH conditions. Since the total concentration of carbonic species seldom exceeds a few millimoles per liter, we deal with a system of very low buffer capacity. A few millimoles of acid or base per liter are sufficient to change the hydrogen ion concentration by some orders of magnitude. Thus, heterogeneous chemical equilibria (interaction of the solution with carbonate rocks, cation-exchange reactions with silicate minerals), as well as biochemical processes (CO_2 removal by photosynthesis, CO_2 production by respiration), and the CO_2 cycle between the atmosphere and the natural waters are more significant for H+ ion regulation in natural waters than the buffer contribution of dissolved carbonic species. The dissolved carbonate system is actually a mediator or indicator of the buffer systems of fresh water rather than the sole, or even a principal, buffering agent. [1, 11].

For a water in which the concentration of CO_2 is governed only by an equilibrium between the dissolved carbonate system and the CO_2 of the atmosphere, the following equation describing the interrelation between [H+], [Alk], and partial pressure can be derived:

$$[Alk] = (KK_1 p_{co_2} / [H+])\,(1 + 2\,K_2/[H+]) + K_w/[H+] \qquad (29)$$

Accordingly, for a water in CO_2 equilibrium with the atmosphere, the H+ ion concentration is defined solely by the alkalinity of the water and the partial pressure of CO_2 (i.e., the same pH should ultimately occur for equinormal solutions of NaOH, $NaHCO_3$, or Na_2CO_3). The pH of a solution containing 10^{-3} equivalents of alkalinity per liter in contact with the atmosphere ($p_{CO_2} = 10^{-3.5}$) should have a pH of approximately 8.4.

Therefore, most fresh waters are oversaturated in CO_2 with respect to an equilibrium with the atmosphere. Accordingly, aeration of fresh waters frequently leads to an increase in pH, causing a closer approach to equilibrium conditions. The conclusion that may be drawn from this is that reactions that tend to depress the pH of natural waters, such as ion exchange, $CaCO_3$ deposition, and respiration, in general are kinetically more rapid than the CO_2 exchange with the atmosphere.

The case for the addition or removal of dissolved carbon dioxide will be developed in some detail since the effect of carbon metabolism upon pH causes this case to be of particular interest. Following is a schematic generalized reaction for carbon metabolism.

$$CO_2 + 2\,H_2O \underset{\text{Respiration}}{\overset{\text{Photosynthesis}}{\rightleftharpoons}} (CH_2O)_n + O_2 + H_2O$$

Any addition of $H_2CO_3{}^*$ to a carbonate solution increases both the acidity of the solution and C_T. The alkalinity, unlike the case for the addition of strong acid, remains unchanged however. (Our assumption in this case does not consider any interaction with $CaCO_3$ or precipitation of $CaCO_3$.) The change in C_T, as a result of addition of CO_2 (or removal of CO_2), can be characterized by equation 30.

$$C_T = \frac{[\text{Alk}] - [OH^-] + [H+]}{\alpha_2(2+ [H+]/K_2)} \tag{30}$$

Equation 30 and its graphical representation (Figure 5) are convenient tools for the evaluation of biochemical respiration and CO_2 assimiliation, and for the assessment of metabolic activity from diurnal variations in pH.

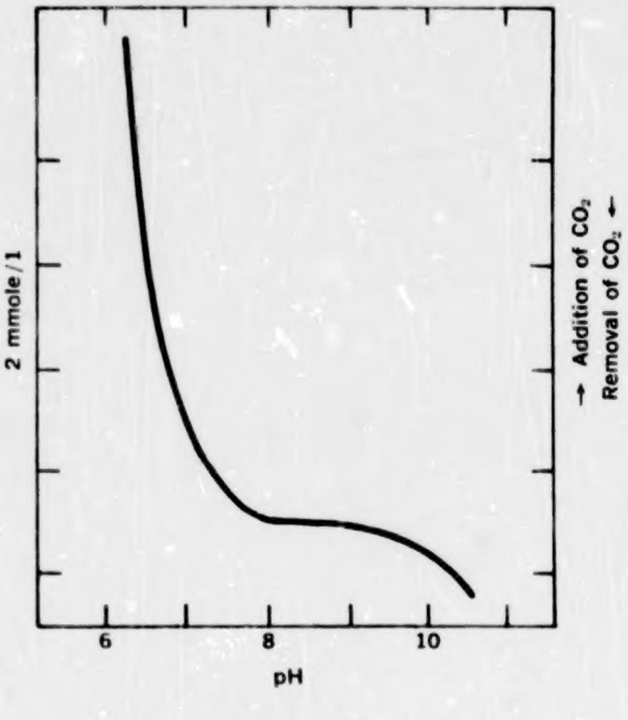


Figure 5 — Addition or Removal of CO_2 to or From a Homogeneous Carbonate Solution (Alk = 10^{-3} = Constant).

ANALYTICAL IMPLICATIONS OF $CaCO_3$ SOLUBILITY EQUILIBRIUM

The $CaCO_3$ solubility equilibrium has been applied in water chemical interpretations for over 50 years. Equations describing the $CaCO_3$ solubility equilibrium (equivalent to equations 15a, b, and c, Table 3) were independently derived by van't Hoff (1890), Tillmans (1912), Kolthoff (1921), Langelier (1936), and others. In many ground waters and a large number of surface waters, the relation between analytically determined concentrations of Ca^{+2}, H^+, and carbonic species (alkalinity or C_T) is in very good accord with the $CaCO_3$ solubility equilibrium. In water technology the same concept has een used with analytical data to predict whether a water will tend to deposit or dissolve aCO$_3$. Figure 6 gives a plot of maximum soluble $[Ca^{+2}]$ as a function of pH for

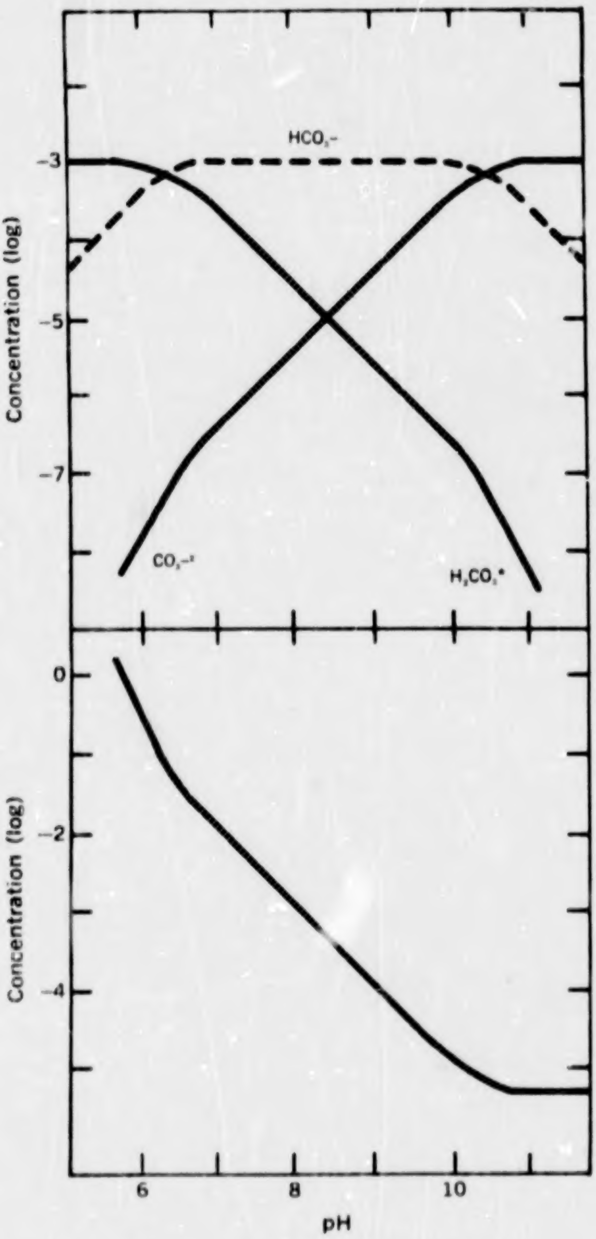


Figure 6 — Dissolved Species of a C_T = 10^{-3}M Carbonate Solution in Equilibrium With $CaCO_3$(s).

$_T = 10^{-3}$M. Equations 15 to 15c can be rearranged in various ways to make them nore suitable for direct use with analytically determinable parameters; for example, if naximum soluble Ca^{+2} is expressed as a function of $[H^+]$ and C_T or $[Alk]$:

$$[Ca^{+2}] = K_s / C_T \alpha_2 \qquad (31)$$

$$[Ca^{+2}] = (K_s/K_2)\ ([H^+]/HCO_3^-]) \qquad (32)$$

or if Ca^{+2} is expressed as a function of $[H_2CO_3^*]$ and [Alk] or C_T:

$$[Ca^{+2}] = (K_sK_1/K_2)\ ([H_2CO_3^*]/\ [HCO_3^-]^2) \tag{33}$$

In both of these expressions, $(HCO_3^-]$ can be substituted for the analytically readily determinable [Alk] or C_T by

$$[HCO_3^-] = ([Alk] - K_w/\ [H^+] + [H^+])\ (1 + 2K_2/\ [H^+]) \tag{34}$$

or

$$[HCO_3^-] = C_T\ \alpha_1 \tag{34a}$$

(At pH values below pH 9, $[HCO_3^-]$ can be set equal to [Alk]; similarly, within the pH range 7 to 9, α_1 is very close to 1.)

Since equations 31 to 33 are conceptually equivalent, one might wonder why 32 is preferentially used in the United States while equation 33 is almost exclusively used in continental Europe. From an operational point of view, equation 33 is analytically more satisfactory for hard, high-alkalinity waters than for soft, low-alkalinity waters (smaller relative error in analytic determination of $[H_2CO_3^*]$ than of $[H^+]$). Correspondingly, the elucidation of "stability" can be rendered more precisely for soft, low-alkalinity waters if based on the analytical determination of $[H^+]$ and [Alk].

It has been suggested especially by Greenwald[12] that Ca^{+2} interacts with HCO_3^- and CO_3^{-2} to form soluble complexes, e.g., $CaHCO_3^+$ and $CaCO_3(aq)$ (see equation 16, Table 3). Although the formation of such complexes is entirely plausible, according to carefully controlled experiments by the author, they do not appear to be of any significance in controlling $CaCO_3$ solubility in the concentrations and pH range found in natural waters. Since $CaCO_3$ solubility is rather dependent on temperature and ionic strength, it is relevant that constants valid at appropriate temperature and proper activity corrections (e.g., those suggested by Larson and Buswell[14]) be used.

OTHER ANIONS

We now add some sulfate, chloride, and nitrate in the form of their sodium or potassium salts. These added ions will have very little influence upon the equilibria already discussed. The solubility product of $CaSO_4$ is of the order of 10^{-5}, so that no $CaSO_4(s)$ will precipitate unless we add more than about 10^{-2} mole of SO_4^{-2} per liter.

PHOSPHATE

The phosphate concentration in natural waters seldom exceeds 0.3 milligram per liter. Upon addition of 10^{-4} mole of phosphate per liter (in the form of Na_2HPO_4), our aluminum and calcium equilibria are influenced. Some of the phosphate will form soluble phosphato-aluminum complexes ($AlHPO_4^+$). The solubility product of $AlPO_4(s)$ is of the order of 10^{-21}, and a calculation will show that $AlPO_4(s)$ will be formed only under slightly acid conditions (pH 5 to 6).

For the interaction of phosphate with calcium we have to consider the following reactions:

$$Ca^{+2} + HPO_4^{-2} = CaHPO_4(s)\ ;\ K = 10^7 \tag{34b}$$

$$5\ Ca^{+2} + 3\ PO_4^{-3} + OH^- = Ca_5(PO_4)_3OH(s)\ ;\ K = 10^{+56} \tag{34c}$$

In the pH range of natural waters, reaction 34c, i.e., the possible formation of hydroxyl-apatite, $Ca_5(PO_4)_3OH(s)$, has to be considered; we might ask ourselves whether some of the added phosphate will convert some of the $CaCO_3$ into apatite. We obtain the equilibrium constant for such a reaction in the following manner:

$$5\,Ca^{++} + 3\,PO_4^{-3} + OH^- = Ca_5(PO_4)_3OH(s); \qquad K = 10^{+58} \quad (34c)$$
$$5\,CaCO_3(s) + 5\,H^+ = 5\,Ca^{++} + 5\,HCO_3^-; \qquad K = 10^{+10} \quad (15b)$$
$$3\,HPO_4^{-2} = 3\,H^+ + 3\,PO_4^{-3}; \qquad K = 10^{-36} \quad (34d)$$
$$H_2O = H^+ + OH^-; \qquad K = 10^{-14}$$

$$5\,CaCO_3(s) + H^+ + 3\,HPO_4^{-2} + H_2O = Ca_5(PO_4)_3OH(s) + 5\,HCO_3^-; \qquad K = 10^{+18} \quad (34e)$$

If we now compute the free energy, ΔF, for the conversion of $CaCO_3$ into apatite by means of the equation, $\Delta F = RT\ln\frac{Q}{K}$, where Q is the quotient of the reactants. In order to compute Q, we assume the following values for the reactants: $[H^+] = 10^{-8}$, $[HCO_3^-] = 10^{-3}$, $[HPO_4^-] = 10^{-6}$. Then we obtain

$$\Delta F = RT\ln\frac{[HCO_3^-]^5}{[H^+]\,[HPO_4^{-2}]^3\,K_{34e}} = RT\ln(10^{-11}) \qquad (34f)$$

corresponding to a ΔF of approximately −15 Kcal, i.e., reaction 34e will proceed from left to right until a new state of equilibrium is reached. At pH 7, the total amount of phosphorus in equilibrium with hydroxyl-apatite is of the order of 10^{-6}M (0.03 mgP/l). Of course, such a figure is only approximate since the constants applied are not known with good precision, but the tentative result suggests that at the sediment and water interface the phosphorus concentration will be buffered by the presence of hydroxyl-apatite as a stable solid phase. This conclusion, if verified, is of utmost significance in connection with the eutrophication of lakes, because it would suggest that the phosphorus distribution in a lake can be interpreted as a heterogeneous distribution equilibrium between sediments and the lake, i.e., any addition of phosphorus (sewage) would lead to a progressive accumulation of phosphorus in the sediments.

BIVALENT METAL OXIDES OR CARBONATES

$FeCO_3$

We now add about 0.5 mole of $FeCO_3$ (per liter of solvent) to our system. Much of the iron that occurs in the earth crust is available as Fe(III). But later, when we open our system to atmospheric oxygen, most of the Fe(II) that we have added as $FeCO_3$ will be oxidized to ferric iron. Thus, the solubility conditions we now describe apply for ferrous iron only.

Table 4 — Fe (II) Solubility

Reaction No.	Reaction	log K
35a	$Fe^{++} + 2H_2O = Fe(OH)_2(s) + 2\,H^+$	+12.9
35b	$Fe(OH)_2(s) = Fe^{++} + 2\,OH^-$	−15.1
36	$FeCO_3(s) = Fe^{++} + CO_3^{-2}$	−10.6
37	$Fe^{++} + H_2O = FeOH^+ + H^+$	− 8.3
38	$Fe^{++} + 3\,H_2O = Fe(OH)_3^- + 3\,H^+$	−32
39	$FeS(s) = Fe^{++} + S^{-2}$	−17.4
40	$H_2S(aq) = H^+ + HS^-$	− 7.0
41	$HS^- = H^+ + S^{-2}$	−12.9

Comparison of equation 15a (Table 3) and equation 36 (Table 4) shows that the solubility product of $CaCO_3$ is about 200 times larger than that of $FeCO_3$. Thus, only about 1×10^{-6} to 2×20^{-6} mole of $FeCO_3$ per liter (0.056 to 0.112 mg/l Fe^{+2}) will go into solution without causing any appreciable change in pH or in concentration of carbonate species.

In the pH range of natural waters, soluble bivalent iron consists of Fe^{+2} and $FeOH^{+}$. The solubility of ferrous iron in all carbonate-bearing waters ($C_T > 10^{-4}M$.) within the common pH range (pH 6 to 9) is governed by the solubility product of $FeCO_3$ (equation 36) and not (as is frequently assumed) by the solubility of $Fe(OH)_2$ (equation 35b).

The solubility product constants of $FeCO_3$ and $Fe(OH)_2$ have different dimensions, *i.e.*, mole²/liter² and mole/liter³, respectively; thus, in order to decide which of the solubility products controls Fe(II) solubility, one must evaluate the pH dependence of Fe(II) solubility by using both constants. Figure 7 gives a solubility diagram for $Fe(OH)_2$ in a non-carbonate water; Figure 8 shows maximum soluble Fe(II) for a carbonate-bearing water ($C_T = 10^{-3}M$). A comparison of these two figures shows that the solubility product of $Fe(OH)_2$ governs the solubility of Fe(II) only in waters that contain no carbonate, or are at very high pH. Thus, essentially the same type of equations that have been used quantitatively to describe the solubility relations of $CaCO_3$ can be used to evaluate Fe^{++} solubility in natural waters (substitution of K $FeCO_3$ for K $CaCO_3$). The maximum soluble Fe(II) for a water that is in $CaCO_3$ saturation equilibrium is only about 0.5 percent its calcium content. These considerations apply only up to a pH of about 8 or 9. Above this pH, hydrolysis of Fe^{+2} to $FeOH^{+}$ (reaction 37) will slightly influence the relations for total soluble Fe(II). Up to about pH 10, soluble Fe(II), as a function of [H+] and [Alk] or C_T, can be estimated by means of the following equation:

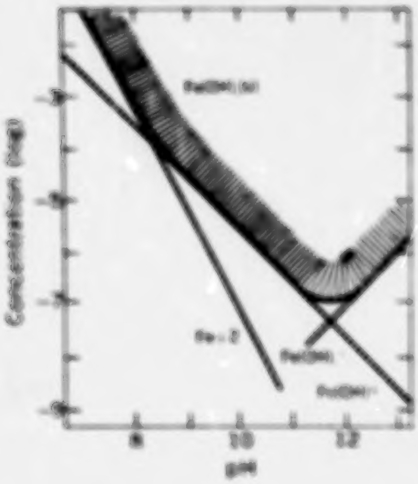


Figure 7 — Solubility of $Fe(OH)_2(s)$ in a Non-Carbonate Solution ($C_T = 0$).

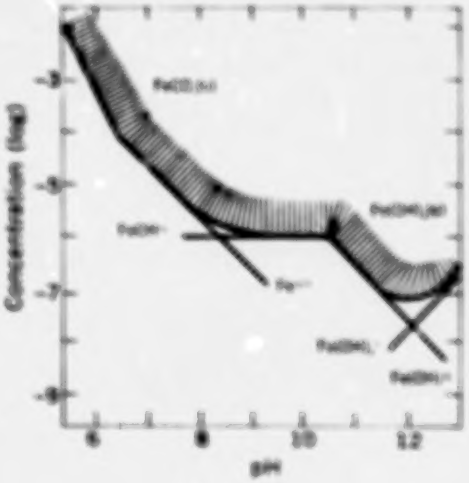


Figure 8 — Maximum Soluble Fe(II) in a Carbonate-Bearing Water ($C_T = 10^{-3}M$) — Only at High pH is the Solubility Controlled by Solubility products of $Fe(OH)_2$.

$$[Fe(II)] = [Fe^{++}] + [FeOH^{+}] = (K_{36}/C_T\alpha_2)\ (1 + K_{37}/[H^{+}]) \qquad (42)$$

In passing, we should be aware that the solubility of ferrous iron can also be controlled by the solubility of ferrous sulfide. The presence of small quantities of S(II) components (H_2S, HS^-, S^{-2} and polysulfides), as they may occur in hypolimnetic waters (e.g., through bacterially mediated reduction of sulfate), is inconsistent with the presence of appreciable amounts of soluble ferrous iron. Frequently, under natural conditions, ferrous iron controls the amount of soluble sulfur, S(II), constituents rather than vice versa. A quantitative evaluation of metal sulfide solubilities is frequently difficult, because solubility products are not known with sufficient accuracy, and the existence of various polysulfide species makes a simple interpretation impossible. As a first approximation, the Fe(II) solubility, as a function of the total sulfide,

$$[S(II)] = [H_2S] + [HS^-] + [S^{-2}],$$

in sulfide-containing waters can be estimated by

$$[Fe^{+2}] = (K_{s0}/[S(II)])(1 + [H^+]/K_{41} + [H^+]^2/K_{40}K_{41}) \qquad (43)$$

Since FeS is less soluble than $FeCO_3$, deposited $FeCO_3$ can be converted by low concentrations of S(II) into black FeS(s) (or FeS_2(s)):

$$FeCO_3(s) + HS^- = FeS(s) + HCO_3^-;\ K = \frac{K_{FeCO_3}K_{HS^-}}{K_{FeS}K_{HCO_3}} = 10^{5.5} \qquad (44)$$

$MnCO_3$

In rocks manganese is less abundant than iron, but like iron it occurs in multiple oxidation states. Upon addition of 0.1 mole of MnO per liter of our solution we would observe that a similarly small quantity of Mn(II) would go into solution. The solubility relations of Mn(II) are very similar to those of Fe(II), as is evident from comparison of Figures 8 and 9. The latter figure has been constructed by using the following equilibrium constants: $\log K_{MnOH_2} = -13$; $\log K_{MnCO_3} = -10.41$; $\log K_1$ ($Mn^{+2} + H_2O = MnOH^+ + H^+$) $= -10.6$; $\log K_{MnHCO_3} +$ ($Mn^{+2} + HCO^- = MnHCO_3^+$)

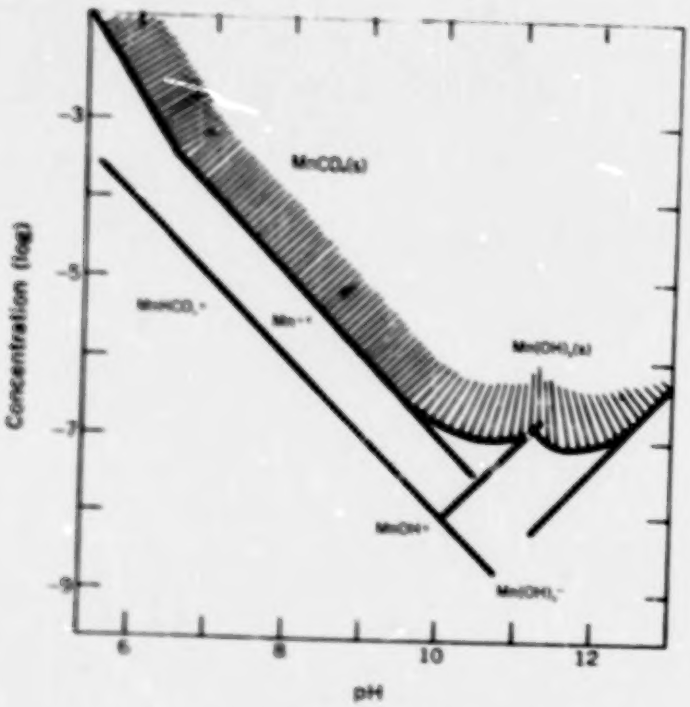


Figure 9 — Maximum Soluble Mn(II) in a $C_T = 10^{-3}$M Water.

= 2. It is seen that the $MnCO_3$ solubility equilibrium controls the solubility of Mn^{+2} of most natural waters. Most of the MnO that has been added to our mixture will be converted to $MnCO_3$. (This conversion will increase the pH of the solution somewhat.)

OTHER METALS

It is beyond the scope of this discussion to estimate solubility equilibrium relations for all the significant cations in water. Metal carbonates do not seem to control the solubility of Mg^{+2} and Cu^{+2}. The solubility of magnesium can be calculated from the solubility product of $Mg(OH)_2$ and from the first hydrolysis constant of Mg^{+2}. The solubility of bivalent copper was estimated in Figure 10 as a function of pH; the following constants were used: $\log K_{Cu(OH)_2} = -18.8$, $\log K_{CuCO_3} = -9.6$ $\log K_1$ ($Cu^{+2} + H_2O = CuOH^+ + H^+$) $= -6.8$. It appears from this figure that the predominant soluble Cu(II) species in most natural waters is $CuOH^+$. The solubility of copper increases again at high pH values, because of hydroxo and/or carbonato complex formation. In the author's laboratory, C. Schneider has determined a stability constant of approximately 10^{10} for a soluble $[Cu(CO_3)_2]^{-2}$ complex.

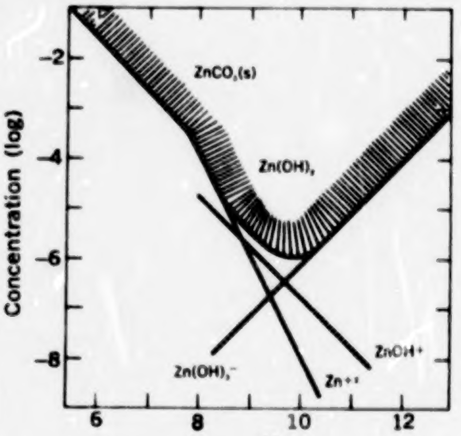


Figure 10 — Maximum Soluble Zn(II) in a $C_T = 10^{-3}M$ Water.

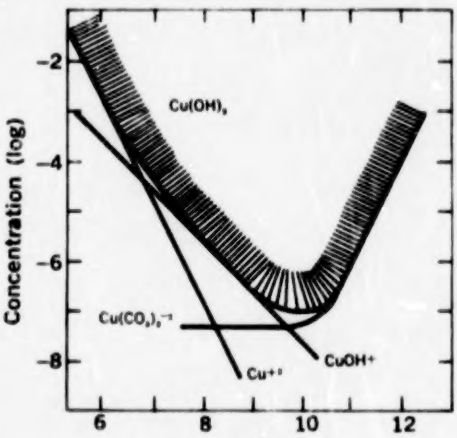


Figure 11 — Maximum Soluble Cu(II) in a $C_T = 10^{-3}M$ Water.

Maximum soluble Zn^{+2} is plotted as a function of pH in Figure 11 ($\log K_{ZnOH_2} = -16.0$; $\log K_{ZnCO_3} = -10.8$, $\log K_1 = -8.7$, and $\log \beta_3$ ($Zn^{+2} + 3\ OH^- = ZnOH_3^-$) $= +14$]. For a C_T of $10^{-3}M$, the zinc solubility is controlled by $ZnCO_3(s)$ below pH 7.5, and by $Zn(OH)_2(s)$ above that pH. With a 10-fold increase of C_T, the zinc solubility would appear to be solely controlled by the $ZnCO_3(s)$ equilibrium.

It might be well at this time to remind the reader that the solubility predictions are based on the selected constants, which might be in error, and all the constants are of course subject to revision as more and better data become available. Frequently, not yet identified species such as soluble $Cu(OH)_2$ or $Zn(OH)_2$ or carbonato complexes might influence the solubility behavior drastically.

OXIDATION REDUCTION POTENTIAL

In our imaginary experiment, we now open the system to the atmosphere representing

GPO 814-105-11

a huge reservoir containing oxygen at a fixed partial pressure, P_{O_2}, of approximately 0.2 atmosphere. The dissolution of oxygen can be described by Henry's law

$$[O_2] = k\, P_{O_2}$$

where k is the equilibrium constant (Henry's law constant) for the oxygen solubility. The oxygen that becomes dissolved might react with some of the constituents in our system. We can visualize for example the oxidation of ferrous iron and Mn(II) to ferric oxide hydrate and to manganese dioxide, respectively:

$$2\,Fe^{+2} + \tfrac{1}{2}\,O_2 + 2\,H_2O = 2\,Fe(OH)_3 + 4\,H^+ \qquad (46)$$

and

$$Mn^{+2} + \tfrac{1}{2}\,O_2 + H_2O = MnO_2 + 2\,H^+ \qquad (47)$$

But any oxygen consumed incipiently in these redox reactions will be refurnished from the atmospheric reservoir so that at equilibrium the dissolved oxygen concentration will still be defined by equation 45. It is thus obvious that in any water system that is in equilibrium with the atmosphere the redox potential is defined by the solubility of oxygen at the given partial pressure of oxygen. All other redox couples, $Fe(OH)_3$ — Fe^{+2}, MnO_2 — Mn^{+2}, etc., will at equilibrium be adjusted in such a way that the ratio of their activities corresponds to the redox potential of the O_2 — H_2O couple. By applying the Peters-Nernst equation, this redox potential of the reaction

$$H_2O = \tfrac{1}{2}\,O_2 + 2\,H^+ + 2\,e^- \qquad (48)$$

can be defined by

$$E_H = E_o + \frac{RT}{nF}\,\ln\,\frac{[H^+]^2\,P_{O_2}^{\;1/2}}{[H_2O]} \qquad (49)$$

or for conditions in the model

$$E_H = 1.23 + \frac{0.059}{2}\,\log\,\frac{[H^+]^2\,P_{O_2}^{\;1/2}}{1} \qquad (49b)$$

where E_H is the electrode potential of the half reaction (48), as compared with the standard hydrogen gas — hydrogen ion couple:

$$H_2 + 2\,H^+ + 2\,e,\ E_o = 0 \qquad (50)$$

By this convention, the potential is negative if the reductant in the half reaction under consideration is a better reductant than hydrogen gas and should reduce H^+ to H_2; in other words, a high E_H can be interpreted as a high oxidation intensity, or more precisely, a low electron activity, and a low (negative) E_H reflects a high electron activity.

On the basis of equation 48, the E_H of our system is given by

$$E_H = 1.23 + 0.0295 \times \tfrac{1}{2}\log 0.2 - 0.059\ pH \qquad (51)$$

or

$$E_{H\,(pH\,6)} = 0.773\ v;\quad E_{H\,(pH\,7)} = 0.714\ v;\quad E_{H\,(pH\,8)} = 0.665\ v$$

These values show that E_H is slightly dependent on pH. The presence of dissolved oxygen is certainly a dominant factor in the oxidation intensity of a water, but it is interesting to note that the potential is remarkably insensitive to changes in the dissolved oxygen concentration. Reducing the oxygen concentration 99 percent, i.e., from 10 to 0.1 milligram per liter, will lower the potential by only 30 millivolts.

THE ABUSE OF E_H MEASUREMENTS IN NATURAL WATERS

Under certain circumstances, electrode potentials can be determined experimentally by inserting an inert metal like platinum in combination with a reference electrode into the solution. With the availability of such an experimental method for electrode potential measurements, it becomes very tempting to use such a procedure for the investigation of reduction and oxidation conditions in waters. For nearly 40 years sanitary engineers and water chemists have based results on the misconception that they were able to evaluate the total oxidative (reductive) capacity as well as oxidation (reduction) intensity by such a comprehensive technique. Unfortunately, these measurements have, in the opinion of the author, failed to yield results amenable to intelligible interpretation. Similarly, anyone who attempts to verify equation 51 for the oxygen-water system by measuring the E_H in an oxygen equilibrated water soon becomes frustrated by the significant discrepancy between observed and calculated data, and by his failure to obtain reproducible E_H readings.

CONCEPT VERSUS MEASUREMENT OF THE POTENTIAL

In textbooks, generally, the concept of electrode potential in oxidation reduction processes is introduced by considering the thermodynamic properties of electrochemical cells. It is necessary however to distinguish between the concept of the potential, as it is employed by Latimer[3] and others, and the measurement of an electromotive force in an actual cell. Potentials quoted by Latimer and by others have been derived from equilibrium data, thermal data, and the chemical behavior of a couple with respect to known oxidizing and reducing agents, and from the direct measurements of cells. The conceptual meaning of a particular potential, in the thermodynamic sense, is that it is the equivalent free energy, i.e., the free energy change per mole of electrons associated with a given oxidation or reduction:

$$E_H = \frac{\Delta F}{nf} \qquad (52)$$

where ΔF is the free energy, f is the Faraday, and n is the number of moles of electrons involved per mole of reactant. There is no a priori reason to identify the thermodynamic potentials with measurable electrode potentials in a given aqueous system.

The measurement of an electrode potential involves a question as to the electrochemical reversibility or irreversibility of the electrode reaction characterized by the rate of electron exchange at the electrode (exchange current). It is realized that E_H measurements are of great value in a few systems for which the variables are known and under control.

Some of the essential principles involved in the measurement of an electrode potential can be qualitatively described by a consideration of the behavior of a single electrode (platinum) immersed into a $Fe^{+2} - Fe^{+3}$ solution. To cause the passage of a finite current at this electrode, it is necessary to shift the potential from its equilibrium value. One thus obtains a curve depicting the electrode potential as a function of the applied current (polarization curve). At the equilibrium potential, i.e., at the point of zero applied current, the half reaction

$$Fe^{+3} + e \rightleftarrows Fe^{+2}$$

is at equilibrium; but the two opposing processes, the reduction of Fe^{+3} and the oxidation of Fe^{+2} proceed at an equal and finite rate that can be expressed by the exchange current. As indicated in Figure 12, the net current can be visualized as the algebraic summation of two opposing currents (cathodic and anodic). The rate of Fe^{+3} reduction (conventionally expressed as cathodic current) generally increases exponentially with more negative electrode potential values and is furthermore a function of the concentration of Fe^{+3} and of the effective electrode area. Similar considerations apply to the rate of Fe^{+2} oxidation (anodic current), which is proportional to $[Fe^{+2}]$, electrode area, and the exponential of the potential. It is obvious from the schematic representation of Figure 12 that in the case of $Fe^{+2} - Fe^{+3}$, provided that the concentration of these ions is sufficiently large, e.g., 10^{-3} to 10^{-4}M, an infinitesimal shift of the electrode potential from its equilibrium value will make the half reaction proceed in either of the two opposing reactions. Operationally, the measurement of the equilibrium electrode potential in such a case is feasible. We might contrast such behavior with the conditions we would encounter in attempting the measurement of the electrode potential in distilled water containing dissolved oxygen. A schematic representation of the polarization curve for this case is given in Figure 13. The equilibrium electrode potential should

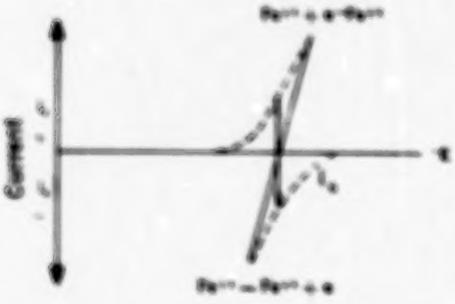


Figure 12 — E_H Measurement in FE^{+2} — Fe^{+3} System.

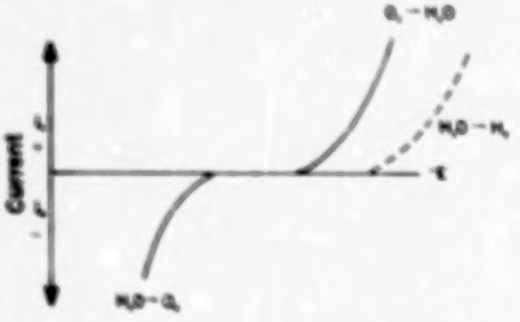


Figure 13 — E_H Measurement in H_2O-O_2 System.

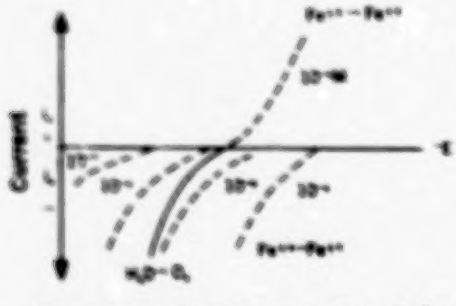


Figure 14 — E_H Measurement in Fe^{+2} — Fe^{+3} System — Occurence of Mixed Potential Because of Low Concentration of Fe^{+2}.

again be located at the point where the net applied current (i.e., the algebraic sum of cathodic and anodic currents) is zero. The exact location is rendered very difficult. Over a considerable span of electrode potentials, the net current is virtually zero; similarly, the electron exchange rate, or the exchange current reflecting the opposing rates of the half reaction,

$$H_2O \rightleftharpoons \frac{1}{2}\, O_2 + 2\,H^+ + 2\,e$$

is virtually zero. Operationally, a remarkable potential shift must be made to produce a finite net current and the current drawn in the potentiometric measurement is very large compared with the exchange current. Even with modern instrumentation in which the current drain can be made extremely low, the experimental location of the equilibrium potential is ambiguous. Furthermore, because of the negligible exchange current, the rate of attainment of the equilibrium potential is very low. The measured potentials drift for hours or even days, and the steady state potential, which is essentially reached, is neither reproducible nor indicative of the thermodynamic electrode potential. Such a system is called an electrochemical irreversible system, and its redox reactants are called non-electroactive. Many redox reactants encountered in natural waters behave

irreversibly at inert electrodes; these reactants include sufide-sulfur-sulfite-sulfate, NO^{3-} — NO_2^- — N — NH_2OH —NH_3, ClO_3^- —OCl^- —Cl^-, most organic redox couples, etc.

The measuring electrode is very easily contaminated by insidious trace quantities of tensioactive materials. Although such a contamination does not necessarily affect the equilibrium position of the potential, it generally leads through adsorption to a significant reduction in effective electrode area and thus reduces markedly the exchange current, which in turn results in a much more sluggish response of the electrode; thus systems that are otherwise electrochemically reversible may become irreversible.

It is necessary to introduce an additional and possibly most important restriction regarding the measurement of E_H: The point of zero-applied current in the polarization curve is not necessarily the equilibrium potential. Figure 14 schematically depicts the polarization curves for the electroactive Fe^{+2} — Fe^{+3} system at various concentrations of Fe^{+2} and Fe^{+3}. The measured equilibrium electrode potential is in accord with the potential calculated according to the Peters-Nernst equation as long as the concentrations of Fe^{+2} and Fe^{+3} are larger than about $10^{-5}M$. (This threshold concentration depends on the effective electrode area.) Below these concentrations, the measured potential can no longer be interpreted in terms of the Peters-Nernst equation. If for example, Fe^{+3} is larger and Fe^{+2} is smaller than $10^{-5}M$, respectively, the measured electrode potential becomes independent of Fe^{+2}.[15] It is evident from Figure 14 that under these conditions the measured equilibrium potential is defined by the point where the equivalent rate of Fe^{+3} reduction is equal to the equivalent rate of H_2O oxidation ($H_2O = \frac{1}{2} O_2 + 2 H^+ + 2 e$). Such a potential is of course no longer characteristic of the Fe^{+3} — Fe^{+2} system. Such a potential is called a mixed potential and bears no simple relationship to the activities of the reacting species. Correspondingly, in a solution of Fe^{+2} containing less than $10^{-5}M$ Fe^{+3}, the measured potential would drift to a value where the rate of reduction of H^+ (or H_2O) would just equal the rate of oxidation of Fe^{+2} to Fe^{+3}. Since Fe^{+3} ions are produced in the reaction, the measured potential would slowly be shifted until eventually an equilibrium would be reached (e.g., after days) in which both half reactions would be at the same potential. This potential would have no bearing, however, on the incipient activities of Fe^{+2} and Fe^{+3}. It is obvious that minute trace quantities of oxidants other than H^+, e.g., oxygen or an oxide film at the electrode, that are reduced at less negative potentials than H^+ might significantly affect the potentiometric reading.

Of the redox reactants in natural water systems, Fe^{+2} and Fe^{+3} may be among the most electroactive species. As is evident from the solubility considerations given in this paper for Fe^{+2} and Fe^{+3}, the concentration of free Fe^{+2} and Fe^{+3} should very seldom exceed $10^{-5}M$.

We must, therefore, conclude that most E_H measurements carried out in natural water systems represent mixed potentials. Under conditions of a mixed potential, a net chemical reaction is proceeding at the electrode and the potential is not characteristic of either half reaction. The measured value of the electrode potential cannot be interpreted quantitatively by simple relationships such as that given by the Peters-Nernst equation.

It certainly might be expected that fresh waters containing primarily oxidizing agents give high E_H measurements and those containing predominantly reducing agents exhibit low E_H readings, but a quantitative interpretation does not appear to be justified. A trustworthy analysis of some of the pertinent constituents of the water, e.g., O_2, HS^-, NO_2^-, NH_4^-, Fe(II), and Fe(III), that can be carried out more precisely and usually faster and simpler than an E_H measurement is generally much more informative than an E_H reading.

FE(III) AND Mn(IV)

Fe(III)

After this digression into a discussion of the concept and measurement of the ORP, we resume the discussion of aqueous iron. In oxygenated water, ferrous iron is oxidized to the ferric iron. The solubility of Fe(III) in natural waters is controlled by the solubility of ferric hydroxide or ferric oxide hydroxide, FeOOH. The equilibrium constants (reactions 53-57) used in the construction of Figure 15 are listed in Table 5.

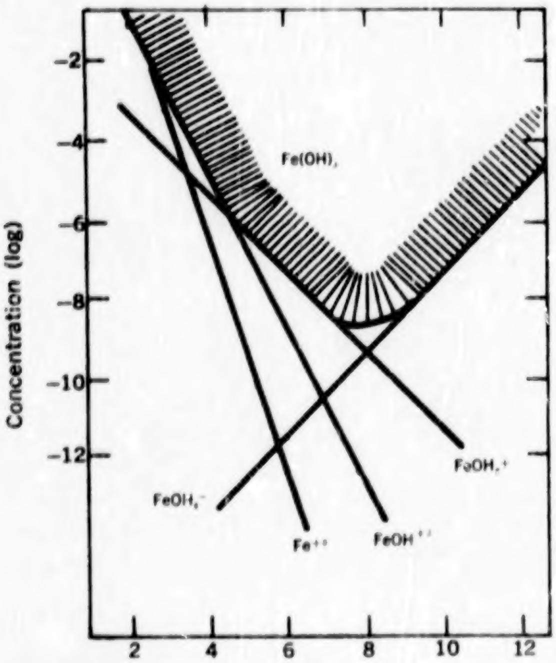


Figure 15 — Maximum Soluble Fe(III).

Table 5 — Fe (III) Solubility

Reaction No.	Reaction	log K
53	$Fe(OH)_3(s) = Fe^{+3} + 3\,OH^-$	—36.0
54	$Fe(OH)_3(s) = FeOH_2^+ + OH^-$	—14.77
55	$Fe\,(OH)_3(s) = FeOH^{+2} + 2\,OH$	—24.17
56	$Fe(OH)_3(s) + OH^- = FeOH_4^-$	— 5.0
57	$FePO_4(s) = Fe^{+3} + PO_4^{-3}$	—23.0 [16]
58	$Fe^{+3} + HPO_4^{-2} = FeHPO_4$	+ 8.4 [16]
59	$Fe^{+3} + SiO(OH)_3^- = FeSiO(OH)_3^{+2}$	+ 9.3 [16b]

According to Figure 14, at pH 7 the following constituents of soluble Fe(III) are in saturation equiiibrium with $Fe(OH)_3(s)$: $Fe^{+3} = 10^{-15}$; $FeOH^{+2} = 6 \times 10^{-11}$;

$FeOH_2^+ = 2 \times 10^{-8}$; $FeOH_4^- = 10^{-12}$. Total soluble Fe(III) is thus in the order of only 1 microgram per liter.

For an air-saturated water with an E_H of 0.717 volts (pH 7), the equilibrium concentration of Fe^{+2} can be calculated by applying the Peters-Nernst Equation to the reaction

$$Fe^{+2} + 3\,OH^- = Fe(OH)_3(s) + e;\ E_0 = -1.31\ v \qquad (60)$$

$$0.717 = -1.31 + 0.059 \log \frac{1}{[Fe^{+2}]\,[OH^-]^3} \qquad (61)$$

For pH 7, the calculated equilibrium concentration of Fe^{+2} amounts to approximately 5×10^{-14}.

Virtually no iron should exist in solution in equilibrium with the atmosphere. This does not appear to be in accord with the analytical findings for real systems. Real systems may not, however, be in equilibrium with oxygen.[18] Furthermore, the solubility of ferric iron might be enhanced by complex formation with inorganic constituents, e.g., phosphato and silicato complexes, or organic constituents. Analytically, it is rather difficult to distinguish between dissolved and suspended iron. Lengweiler, Buser, and Feitknecht[17] have shown that with very dilute Fe(III) solutions containing Fe^{59} as tracer and brought to a pH between approximately 5 and 12 all the iron hydroxide can be sedimented by ultracentrifugation (93,000 g, 180 min). The size of the $Fe(OH)_3$ particles varies with the pH of the solution. The diameter can be as small as 100 A°. It is obvious that filtration (even through membrane filters) does not always provide a satisfactory operation for the distinction between the dissolved and suspended fractions of a particular species.

As we have seen, both ferrous and ferric iron generally are not very soluble in natural waters. Despite this low solubility, the capability of iron to undergo reversible oxidation and reduction reactions plays a significant role in the chemistry and biology of natural waters. In limnology, the redox reactions of iron are related to the metabolic cycles of nearly all other important elements and to the distribution of oxygen in a body of water.[18] During the seasonal variations in an eutrophic lake, the continuous sequence of circulation and stagnation is accompanied by oxidation and reduction as well as precipitation and dissolution of iron. This leads to a progressive accumulation of iron in the lake sediments. In many lakes, interesting correlations between the concentrations of Fe(II) and Fe(III) and those of phosphates and silicates are observed. The strong affinity of phosphates and silicates to Fe(III) (reactions 57-59) might provide an important clue for a more quantitative interpretation of such correlations.

MANGANESE

Figure 16 gives a redox potential pH diagram for manganese. At the potential of an air-saturated solution Mn^{+2} is thermodynamically unstable. In the absence of strong complex formers, Mn(III) does not occur as a dissolved species. In Figure 16 it is seen that $MnO_2(s)$ is the only manganese oxide phase that would be stable in oxygenated waters. In deep-sea sediments MnO_2 is indeed an abundant constituent. The manganese concentration found in aerated fresh waters probably consists of a mixture of Mn(II) and colloidally dispersed MnO_2. The oxidation of Mn(II) to higher valent manganese oxides has been found[20] to be strongly pH dependent and autocatalytic. Below pH 8.5, the rate of oxygenation is extremely low. The oxygenation does not lead to stoichiometric oxidation products such as MnO_2, MnOOH, or Mn_3O_4. The results of studies on the oxidation of Mn(II) can best be interpreted by assuming that the oxidation products consist of MnO_2 onto which various quantities of Mn^{+2} have been adsorbed. Colloidal aqueous manganese dioxide has been shown to have a remarkable ion-exchange capacity for Mn(II) and other metal ions, since this property is strongly dependent upon pH. Sorption capacities in excess of 0.5 mole of Mn^{+2} per mole of MnO_2 are found in the slightly alkaline pH range.[21]

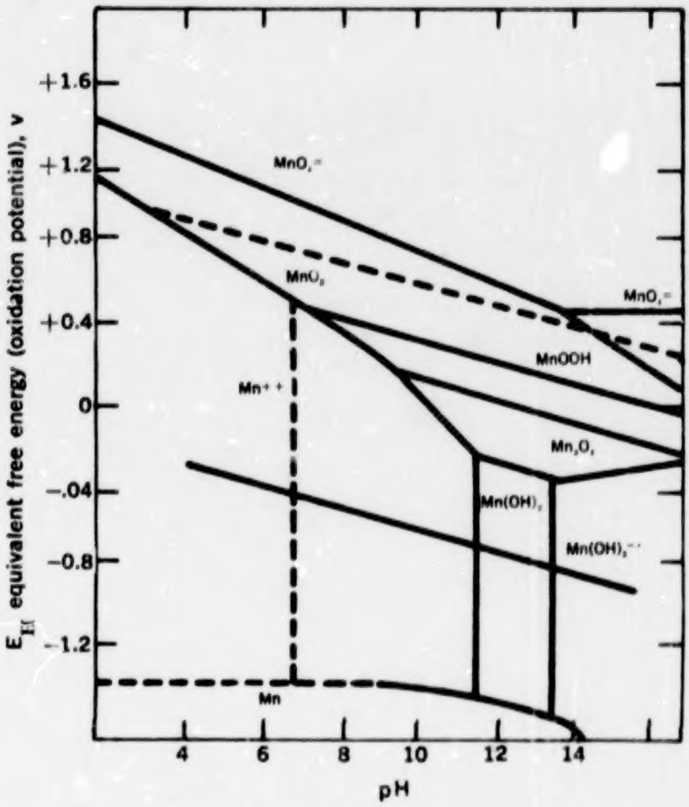


Figure 16 — Redox Potential — pH Diagram for Aqueous Manganese.

Aqueous manganese dioxide, to a pronounced degree, possesses some of the characteristics that appear to be generally applicable to an interpretation of properties of polyvalent metal oxide hydrates. In a similar way, ferric hydroxide exhibits cation-exchange properties, especially at high pH. At high pH values, exchange capacities as high as 1 equivalent per mole of hydrous metal oxide (e.g., Mn^{+2} on $Fe(OH)_3$) are not uncommon. Cation exchange on the hydrous oxides is comparable to the cation exchange on clay materials. Such ion-exchange phenomena on hydrous metal oxides and other precipatates (solid solutions) represent special cases of heterogeneous metal ion buffers. The concept of solid solutions provides one possible explanation for the observed occurrence of certain impurities (e.g., metal ions) in sediments that have subsided from solutions apparently (without considering the activity coefficients of the solid) unsaturated with respect to the impurity.

FINAL REMARKS

The imaginary experiment could of course be continued at great length and many of the cases that have been discussed should be treated in much more detail. But the primary aim of this discussion was to show the simple methodological tools that the chemist can use to arrive at conclusions on mineral relations in natural waters. All the information gathered together in the examples discussed has been taken from standard reference tables on the energies or on the relative stabilities of various compounds. It is regrettable that this easily available information has not been sufficiently used in the past to help answer many of the qualitative and quantitative questions involved in the mineral relations of natural waters and to serve as a guide in the interpretation of analytical results.

An attempt has been made to describe the stability relationships of the distribution of the various soluble and insoluble forms through rather simple graphic representations. The principle involved in elucidating the equilibrium relationships consists essentially in writing down as many equations as one has unknowns and to solve them. A simultaneous graphical representation of all the requisite equations gives the means for attacking even very complicated systems. Two types of graphical treatments have been used in this discussion: first, equilibria between chemical species in a particular oxidation state as a function of pH and solution composition; second, the stability of different oxidation states (potential-pH diagram) as a function of pH and solution composition. Diagram of the latter type require for their delineation an intensity factor as a variable representing the stability of the various oxidation states. Since the redox potential, in most cases, cannot be measured operationally (it can be computed, of course), the potential-pH diagrams are somewhat less amenable to a simple and direct interpretation than the log concentration pH diagrams.

From the few examples discussed, it has become apparent that there are considerable gaps in our information. Many equilibrium constants are only approximately known and some are missing. But we also lack information on the real systems. Many reported analytical data of natural waters are unreliable. For example, the author would doubt the reliability of most of the results that have been published on the Fe(III) and Fe(II) content or on the phosphorus and sulfide concentration of natural waters. Many of the analytical methods we use are not sufficiently specific; it is also difficult to distinguish analytically between dissolved and suspended species. Extensive redox potential measurements in natural media have failed to yield information that can be interpreted quantitatively. It is hoped that all these obvious shortcomings represent an incentive for careful investigations in the future.

REFERENCES

1. Sillen, L. G., Proc. Inter. Ocean. Congress, Publ. No. 67, p 549, AAAS, Washington, D. C., 1961
2. Bjerrum, J., G. Schwarzenbach and L. G. Sillen, "Stability Constants," The Chem. Soc., London, 1958
3. Latimer, W. M., "Oxidation States," Prentice Hall, 1952
4. Goldschmidt, V. M., J. Chem. Soc., 655 (1937)
5. Cited from Hem, J. D., "Geological Survey Water-Supply Paper 1473" (1959)
6. Lagerstrom, G., Acta Chem. Scand., *13*, 722 (1959)
7. Weber, W. J. and W. Stumm, J. Chem. Engr. Data, July 1963
8. Brosset, C., G. Biedermann and L. G. Sillen, Acta Chem. Scand., *8*, 1917 (1954)
9. Matijevic, E., et. al., J. Phys. Chem., *65*, 826 (1961)

10. Sillen, L. G., in Treatise on Analytical Chemistry, Part 1, Vol. 1, p. 277, Interscience, New York (1959)

11. Weber, W. J., and W. Stumm, Jour. A.W.W.A., *55*, Oct. 1963

12. Greenwald, I., J. Biol. Chem., *141*, 789 (1941)

13. Larson, T. E., and A. M. Buswell, Jour. A.W.W.A., *34*, 1667 (1942)

14. Morgan, J. J., Thesis, Harvard University, 1963

15. Coursier, J., Anal. Chim. Acta, *7*, 77 (1952)

16b. Weber, W. J. and W. Stumm, to be published (1964)

17. Lengweiler, H., W. Buser and W. Feitknecht, Helv. Chim. Acta, *44*, 805 (1961)

18. Stumm, W., and G. F. Lee, Ind. Eng. Chem., *53*, 143 (1961)

19. Stumm, W., and G. F. Lee, Schweiz. Z. Hydrologie, *22*, 295 (1960)

20. Morgan, J. J., and W. Stumm, Presented ACS Meeting, New York, Sept. 1963

21. Morgan, J. J., and W. Stumm, J. Coll. Sci., *19*, 347 (1964)

Dr. Gerard A. Rohlich
Professor of Sanitary Engineering
University of Wisconsin, Madison

SUMMARY

The analysis of water measurement data for basic relationships among hydrological, chemical, and biological parameters is discussed. The data to be assembled and interpreted by the sanitary engineer concerned with environmental problems usually are gathered by scientists in other disciplines; thus, the sanitary engineer must rely heavily on the validity of their interpretations. A properly planned program with well-defined objectives is of paramount importance. Inadequate planning of sampling procedures is more likely to lead to erroneous conclusions than are correlation and statistical handling of data. Another source of error in interpreting results and drawing conclusions in the study of water supply and water pollution control problems lies in the relating of laboratory studies to field situations. The dynamic system in nature is frequently oversimplified; adjustments of variables in laboratory experiments seldom parallel changes in the natural system.

DATA INTERPRETATION — DRAWING CONCLUSIONS

In giving consideration to the subject of our discussion, "Data Interpretation — Drawing Conclusions," I was reminded of the story that Professor E. B. Phelps told in the preface to his book *Stream Sanitation*.[1] After pointing out that it might appear that the subject stream sanitation was rather specialized and could not be "contained within definite boundaries such as scientists are so fond of laying down," he then went on to relate that while serving as an expert witness in a stream pollution case he was questioned at length during cross-examination concerning his title of Professor of Sanitary Science and the scope of his expert qualifications. "Are you a biologist?" he was asked, "a chemist? a botanist? Does your knowledge cover the physiology of fish, and the geology of the area?" To all of these questions he felt he had to reply in the affirmative, with qualifications, for his testimony had, in fact, as he states, "trespassed upon all these 'fields' of science."

The sanitary engineer usually finds himself in this position in the interpretation of data and in drawing conclusions, and in fact in his assessment of a situation he must frequently consider many other facets such as flood control, power development, and irrigation, as well as political and economic factors, when practical situations are confronted.

Obviously the engineer relies heavily on the chemist and biologist to supply water measurement data that may be integrated with physical data in order that the overall evaluation can be made.

As has been mentioned on more than one occasion during this conference, there is no substitute for a properly planned program, if this is at all possible, before embarking on an extensive, costly, and time-consuming project. In any research project we are well aware of the need for experimental design in the planning stages, without which the results may not be worth treating statistically.[2] It has been said that to get the right answers we must ask the right questions. Certainly the investigator must ask "what is the objective?" and formulate simple, clear, specific aims that are as refined as possible. If this is done, a good start has been made, providing the objective is realistic in terms of time and resources available. The most important and most difficult task in any program is to know when to stop — the easiest thing to do is to continue to get more data. Frequently, more data are sought in the hope that perhaps by some chance a key piece of information will appear that "may unlock the puzzle and conclusions will fall out and become self-evident." This is usually a forlorn hope.

In the laboratory research experimental method in which an event occurs under known conditions where "as many extraneous influences as possible are eliminated and close observation is possible, relationships between phenomena can be found."[2] The experimental method is not appropriate to all types of research, however. In the field of environmental measurements, as has been emphasized at this meeting, the unknowns and variables remain in many instances unknown; since we do not have the "controlled experiment," we become purely observational investigators. This position is frequently one to which we do not adjust readily. The principles of the experimental method are not to be forgotten, however. The main difference, as Beveridge[2] states, is that the hypotheses are tested by collection of information from phenomena that occur naturally instead of those that are made to take place under experimental conditions. Unfortunately, although considerable useful data are available for formulating conclusions, there are gaps and limitations, and it is unlikely that we shall ever fully understand the ecological pattern involved in man's relationship to the water environment.

Perhaps the biologists are more aware of the complexity of the microcosm with which we are concerned than are some of the rest of us, and, as in the past, will continue to contribute to an understanding of the relationship of the parameters of water quality, which we now know how to measure, to the environmental problem.

Despite the limitations that we may have in understanding the whole structure, there is no escape from the fact that we are in many instances required to obtain results that will have some practical application, initially to aid in understanding a problem, and, through understanding, to arrive at a solution, however inadequate that solution may be in the light of subsequent information. There is, of course, danger in separating our activity from our contemplation. All too often we become too rigid in the cataloging of existing knowledge; in particular those of us in engineering are inclined to rely heavily on mathematical symbols and models (perhaps rather than an understanding of mathematics). As was pointed out by Lord Kelvin, "Nothing can be more fatal to progress than a too-confident reliance on mathematical symbols, for the student is only too apt to take the easier course, and consider the formula, and not the fact, as the physical reality." I don't mean to say that if we have all the data, stored, and readily retrievable we won't be able to find out things that our common sense would not lead us to. But the common sense approach is still useful and should not be discarded.

We must ask ourselves whether the information we would like to have is really going to be useful. If we are considering water quality and water pollution, we must consider carefully the parameters involved in our specific problem. We are inclined to speak glibly about pollution without defining it for the reason that it defies defnition. We recognize that pollution is strictly a relative term and depends upon the particular use a person wants to make of the water. What might be polluted to one user is far from being polluted to another. Consequently, the parameters that might interest one person are quite different from those that might interest another.

If we are to use intelligently the massive accumulation of data and avoid the separation of activity from contemplation, we should refer frequently to the basic considerations of and the reasons for the data gathering. We must ask ourselves critically:

For what purpose do we intend to use these data? What do we wish to find out or define by these data?

As mentioned previously, the engineer concerned with environmental problems usually is confronted with the assembly and interpretation of data gathered by scientists in other disciplines. In pollution studies he must rely heavily on the chemist, biologist, hydrologist, meteorologist, oceanographer, economist, and frequently on the political scientist and lawyer before his final conclusions can be made. The engineer, like the others, can not claim to be an expert in all these fields and so must rely heavily on the validity of the individual expert's interpretations of his data. Frequently, there is an imbalance in the kinds of data obtained, and the mistake of placing reliance on meager data is a pitfall to be avoided. Although precision my be apparent with minimal information, extension of the study may well indicate that the limited data at best were reflecting a low or high portion of a trend that was in fact related to some other environmental factor that may or may not have been properly considered.

Thus, in the interpretation of data, the qualitative factor should be evaluated before the quantitative aspects are considered. The extent and replication of sampling in relation to the complexity of the area under study are of obvious importance as guides in determining the reliability of the conclusions drawn. In dynamic systems such as lakes and streams, physical, chemical, and biological properties are related to time of sampling, and misleading or erroneous conclusions frequently result unless careful consideration is given to the representation of the particular samples. A knowledge of the extent to which homogeneity exists in the body of water sampled is equally as important as the analytical procedures used on the samples obtained. The correlation and statistical handling of data, although not always simple procedures, are much less serious problems and are less likely to lead to erroneous conclusions than the errors resulting from inadequate planning of the sampling procedures. A useful reference in this regard is the Geological Survey Water-Supply Paper 1473 on the study and interpretation of the chemical characteristics of natural water.[3]

Another source of error in interpretating results and drawing conclusions in the field of water supply and water pollution control lies in the relating of laboratory studies to field situations. The relation of occurrences in the laboratory to the dynamic system in nature is frequently oversimplified, and the conclusions drawn from adjustment of variables in laboratory experiments seldom parallel similar changes that can be made or might occur in the field.

Dr. Stumm has pointed to the fact that many gaps remain in our information regarding the chemistry of natural waters in relation to water quality and Dr. Hynes has made reference to the complexity of interpretation of biological data with reference to water quality. Although there seems to be little need to emphasize the statements of these experts in chemistry and biology, their papers serve as reminders of the dangers of oversimplification in interpreting data and in drawing conclusions.

REFERENCES

1. Phelps, E. B. Stream Sanitation. John Wiley and Sons, Inc., New York, 1944.
2. Beveridge, W. I. B. The Art of Scientific Investigation. 2nd edition. William Heinemann Ltd. London (1953).
3. Hem, J. D. Study and Interpretation of the Chemical Characteristics of Natural Water. Geological Survey Water-Supply Paper 1473. U. S. Government Printing Office, Washington, D.C. 1959.

GPO 814-105-12